PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/00, C12N 15/18, C07K 14/495 — A1

(11) International Publication Number: WO 95/19987

(43) International Publication Date: 27 July 1995 (27.07.95)

(21) International Application Number: PCT/US95/01153

(22) International Filing Date: 25 January 1995 (25.01.95)

(30) Priority Data:
08/188,227 25 January 1994 (25.01.94) US

(71) Applicant: THE SCRIPPS RESEARCH INSTITUTE [US/US]; 10666 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors: LOCKUTOFF, David, J.; 426 Glencrest Drive, Solana Beach, CA 92075 (US). CURRIDEN, Scott, A.; 2432 Caminito Ocean Cove, Cardiff, CA 92007 (US).

(74) Agents: FITTING, Thomas et al.; The Scripps Research Institute, 10666 North Torrey Pines Road, TPC-8, La Jolla, CA 92037 (US).

(81) Designated States: JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: A NEW SENSITIVE METHOD FOR QUANTIFYING ACTIVE TRANSFORMING GROWTH FACTOR-BETA AND COMPOSITIONS THEREFOR

(57) Abstract

The present invention describes a highly sensitive and specific non-radioactive quantitative assay method for quantifying transforming growth factor-beta (TGF-β) in a liquid sample. Also disclosed are TGF-β responsive expression vectors that express the indicator molecule, luciferase, in a dose-dependent response to TGF-β activation. Eucaryotic cells transformed with the disclosed expression vectors are also described. Diagnostic systems in the form of kits for quantifying the amount of TGF-β in a liquid sample using the disclosed methods and expression vectors are described.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

A NEW SENSITIVE METHOD FOR QUANTIFYING ACTIVE TRANSFORMING GROWTH FACTOR-BETA AND COMPOSITIONS THEREFOR

Technical Field

The present invention relates to a sensitive assay method for quantifying the amount of active transforming growth factor beta (TGF-ß) and vector compositions for use therein for expressing an indicator molecule in response to TGF-ß activation of a TGF-ß response element in the vector.

Background

Transforming growth factor beta, hereinafter referred to as TGF-ß, is a 25 kilodalton (kD) homodimeric protein that belongs to a family of regulators of cell growth and differentiation that includes activins, inhibins, Mullerian inhibiting substance, the Drosophila decapentaplegic complex and bone morphogenic proteins. For review, see, Massague, Ann. Rev. Cell Biol., 6:597-641 (1990); Roberts et al., In Peptide Growth Factors and Their Receptors, Sporn et al., Eds., Springer-Verlag, Berlin, 1:419-472 (1990); and Hoffman, Curr. Opin. Cell Biol., 3:947-952 (1991). TGF-ß was initially defined by its ability to induce morphological transformation of fibroblastic cells in monolayer culture and stimulation of colony formation in soft agar. Delarco et al., Proc. Natl. Acad. Sci., USA, 75:4001-4005 (1978) and Todaro et al., Proc. Natl. Acad. Sci., USA, 77:5258-5262 (1980).

Three distinct molecular isoforms of TGF-ß, the genes of which are located on different chromosomes, have been identified in mammals and are designated TGF-ß1, TGF-ß2 and TGF-ß3. Derynck et al., Nature, 316:701-705 (1985); Hanks et al., Proc. Natl. Acad. Sci., USA, 85:71-72 (1988); and Madisen et al., DNA, 7:1-8 (1988). Each of the isoforms are first synthesized as high molecular weight latent or inactive precursor polypeptides that are then processed to 12.5 kD

monomers. Activation of the latent complex can occur through a variety of physiochemical or enzymatic treatments as well as in various tissue culture systems. For review, see Barnard et al., Biochim. Biophys. Acta., 1032:79-87 (1990). Two processed monomers then dimerize to form biologically active TGF-ß.

The activation process must occur to allow binding of the dimerized TGF-ß to the high affinity TGF-ß receptors expressed on the surfaces of all normal cells and most all neoplastic cells. Tucker et al., Proc. Natl. Acad. Sci., USA, 81:6757-6761 (1984); Frolik et al., J. Biol. Chem., 259:10995-11000 (1984); Pircher et al., Biochem. Biophys. Res. Commun., 136:30-37 (1986).

Although some TGF-ß activation systems generate the mature TGF-ß in nanogram quantities, the majority liberate picogram amounts. These low concentrations, however, are sufficient to induce a variety of biological responses such as macrophage chemotaxis (Wahl et al., Proc. Natl. Acad. Sci., USA, 84:5788-5792 (1987)), inhibition of endothelial cell migration and proliferation (Heimark et al., Science, 233:1078-1080 (1986)), stimulation of extracellular matrix deposition (Ignotz et al., J. Biol. Chem., 261:4337-4345 (1986)) and decreased plasminogen activator (PA) activity as a result of decreased PA production (Laiho et al., J. Cell. Biol., 103:2403-2410 (1986) and Flaumenhaft et al., J. Cell. Physiol., 152:48-55 (1992)) along with increased secretion of its inhibitor, plasminogen activator inhibitor-1 (PAI-1) (Laiho et al., J. Biol. Chem., 262:17467-17474 (1987)).

PAI-1 is the primary inhibitor of both tissue-type plasminogen activator (t-PA) and urokinase-type plasminogen activator (u-PA), and as such is a potent anti-fibrinolytic molecule. PAI-1 synthesis by cultured cells in vitro is induced by a variety of molecules including cytokines, growth factors, hormones, and other agents such as endotoxin and phorbol myristate acetate. Nuclear transcription run-on assays demonstrate that the regulation of PAI-1 by many of these

agents, including TGF-ß, occurs primarily at the level of transcription.

TGF-ß released from platelets may be an important negative regulator of the fibrinolytic system of the vessel wall since the TGF-ß in releasates of thrombin-activated platelets causes large increases in PAI-1 synthesis by endothelial cells. This increased PAI-1 synthesis may account for the resistance of platelet-rich thrombi to thrombolytic therapy. The accumulation of PAI-1 in the extracellular matrix in response to TGF-ß protects matrix proteins from proteolytic degradation. Thus, the induction of PAI-1 by TGF-ß may also play a role in both wound healing and fibrotic responses.

These and other biological effects of TGF-ß activity have been used to develop a variety of semiquantitative and quantitative bioassays including those based on chondrogenesis, inhibition of DNA synthesis and cell growth, differentiation, migration or PA activity. Differentiation-based assays include the induction of cartilage specific proteoglycan expression (ED_{50} = 5 ng/ml; 200 pM) (Ogawa et al., In Peptide Growth Factors, Barnes et al., Eds, Academic Press Inc., 198:317-327 (1991); Seyedin et al., Proc. Natl. Acad. Sci., USA, 82:2267-2271 (1985)) and inhibition of rat L6 myoblast differentiation (ED_{50} = 0.2 ng/ml; 8 pM) (Florini et al., J. Biol. Chem., 261:16509-16513 (1986)). An ED_{50} represents the half-maximal amount of factor required to produce an effect, activation or inhibition, on differentiation of target cells. The abbreviations ng/ml, pg/ml, nM and pM respectively stand for nanograms/milliliter, picograms/milliliter, nanomolar and picomolar. These assays are utilized primarily for studying differentiation rather than for quantification of TGF-ß.

Assays based on TGF-ß's ability to inhibit DNA synthesis and cell growth in mink lung epithelial cells (MLE cells) (ED_{50} = 10-20 pg/ml; 0.4-0.8 pM) (Lucas et al., In Peptide Growth Factors, Barnes et al., Eds, Academic Press Inc. 198:303-316 (1991) and Danielpour et al., J. Cell. Physiol., 138:79-86

(1989)), African green monkey kidney epithelial cells (ED_{50} = 1 ng/ml; 40 pM) (Holley et al., Proc. Natl. Acad. Sci., USA, 77:5989-5992 (1980)), rat hepatocytes (ED_{50} = 0.4 ng/ml;16 pM) (Nakamura et al., Biochem. Biophys. Res. Comm., 133:1042-1050 (1985)), and fetal bovine heart endothelial cells (ED_{50} = 75-125 pg/ml; 3-5 pM) (Qian et al., Proc. Natl. Acad. Sci., USA, 89:6290-6294 (1992)) are sensitive but can be affected by a variety of molecules such as insulin, EGF, PDGF, and bFGF.

Migration and plasminogen activator (PA) activity assays have also been described. The migration of bovine aortic endothelial cells (BAEs) into a denuded area of a monolayer is inhibited by TGF-ß (ED_{50} - 2 µg/ml; 80 pM: sensitivity 10-20 pg/ml; 0.4-0.8 pM) (Sato et al., J. Cell Biol., 107:1199-1205 (1988); Sato et al., J. Cell Biol., 109:309-315 (1989); and Sato et al., J. Cell Biol., 111:757-763 (1990). Migration of BAEs, however, can be simultaneously stimulated by endogenously or exogenously supplied bFGF that can abrogate TGF-ß's inhibitory effect (Sato et al., J. Cell Biol., 107:1199-1205 (1988)). The PA assay for measurement of TGF-ß concentration is very sensitive and rapid (Flaumenhaft et al., J. Cell. Physiol., 152:48-55 (1992)). The assay is based on the ability of TGF-ß to decrease PA activity of BAEs by inhibiting PA synthesis and secretion and by inducing expression of its inhibitor, PAI-1. This assay, however, is also sensitive to other molecules, such as bFGF, that can alter PA activity (Flaumenhaft et al., J. Cell. Physiol., 152:48-55 (1992) and Sato et al., J. Cell Biol., 107:1199-1205 (1988)). The ED_{50} of the assay varies from 1 to 35 pg/ml (0.04-1.4 pM) of TGF-ß depending on differences in basal PA levels and sensitivity to TGF-ß among primary BAE cultures.

The ability of TGF-ß to stimulate PAI-1 expression has recently been used to study TGF-ß receptors. Wrana et al., Cell, 71:1003-1014 (1992) transiently transfected a PAI-1 luciferase construct together with a human type II TGF-ß receptor expression vector into TGF-ß resistant MLE cells.

This luciferase construct contained a short, synthetic TGF-ß response element based on the human PAI-1 promoter and was used to report functional expression of the receptor. Although only used to screen transfected mutant cell lines, this construct appeared to be less sensitive to TGF-ß than the constructs of this invention when transiently transfected into MLE cells, and no information was reported regarding its dose-responsiveness or specificity.

In another study of the TGF-ß-stimulation of PAI-1 expression, Riccio et al., Mol. Cell. Biol., 12:1846-1855 (1992), transiently transfected TGF-ß responsive cells with constructs containing varying regions of the 5'-flanking domain of the human PAI-1 gene to determine the transcription regulatory mechanism used by TGF-ß. All the constructs contained the gene encoding the enzyme chloramphenicol acetyltransferase to provide for an indirect determination of the transcriptional effect of the various constructs. With this approach, a 67 base pair region that contained binding sites for the two proteins, CCAAT-binding transcription factor-nuclear family I family and USF factor. Both sites were necessary to obtain TGF-ß induction. The constructs, however, were not utilized in assays to determine dose-responsiveness nor measure the amount of TGF-ß in a sample.

The most specific assays for TGF-ß are the radioreceptor, radioimmumoassay (RIA), and enzyme-linked immunosorbent assay (ELISA). Radioreceptor assays using a variety of cell types, such as A549 human lung carcinomas and murine AKR-213, have been described and have ranges of 125 pM/ml to 25 ng/ml (5 pM-1 nM) with ED_{50} of approximately 0.5 ng/ml (20 pM). See, Wakefield et al., J. Cell. Biol., 105:965-975 (1987); Sato et al., J. Cell Biol., 111:757-763 (1990); Lucas et al., In Peptide Growth Factors, Barnes et al., Eds, Academic Press Inc. 198:303-316 (1991) and O'Connor-McCourt et al., J. Biol. Chem., 262:14090-14099 (1987). RIAs specific for TGF-ß1 and ß2 have ED50s of 12 and 37 pM, respectively (Danielpour et al., J. Cell

Physiol., 138:79-86 (1989)). Others, using different antibodies, describe the range of TGF-ß1 specific RIAs to be 6.25-200 ng/ml (0.25-8 nM), with a sensitivity of 2.4 ng/ml (0.1 nM) (Lucas et al., In Peptide Growth Factors, Barnes et al., Eds, Academic Press Inc. 198:303-316 (1991)). As demonstrated by the differences in these results, the affinities of the antibodies can greatly alter the sensitivity of the assay.

Isoform-specific double antibody or sandwich ELISAs (SELISA) are also very sensitive to the affinities of the antibodies. One such assay, using two different monoclonal antibodies specific for TGF-ß1, had a useful range of 0.63 to 40 ng/ml (0.025-16 nM) (Lucas et al., In Peptide Growth Factors, Barnes et al., Eds, Academic Press Inc. 198:303-316 (1991)). Using a combination of isoform-specific turkey and rabbit antibodies, Danielpour et al., J. Cell Physiol., 138:79-86 (1989) created a SELISA with detection limits of 2-5 pg/well (20-50 pg/ml; 0 8-2 pM). Although highly sensitive and specific, SELISAs such as these are not readily available and are expensive.

Although all of these other TGF-ß assays can detect mature TGF-ß, the low concentrations (<2 pM) generated in various biological systems make many of them impractical without prior concentration of the sample. This can result in large losses of the mature growth factor or more importantly activation of latent TGF-ß. Moreover, many of the assays are complicated to establish and can be influenced by other factors present in the samples thus reducing their utility for accurating measuring the amount of TGF-ß in the sample. For this reason, a need exists for a relatively simple, sensitive and nonconfounding assay for TGF-ß.

Brief Description of the Invention

A highly sensitive and specific, non-radioactive assay, for mature (active) TGF-ß has now been developed. When

compared to the sensitive and widely used proliferation-based MLEC method for measuring TGF-ß concentration, the TGF-ß assay method of this invention is more rapid, has comparable sensitivity, and has a greater detection range. Specificity of this novel assay was also higher as evidenced by its relative insensitivity to factors such as EGF and bFGF which can greatly affect other assays. The use of a truncated PAI-1 promoter that does not respond to other growth modulators such as PDGF found in biological samples, the method of this invention can be used in conditions where other bioassays are difficult to interpret. Because of its large range and specificity, the rapid, sensitive, non-radioactive, easily performed assay method of this invention is useful in determining active TGF-ß concentrations in complex solutions.

Thus, the present invention overcomes the limitations of existing methods used to quantify the amount of TGF-ß in a liquid sample. This invention contemplates a method for quantifying the amount of TGF-ß in a sample using a system comprising a TGF-ß responsive cell containing an expression vector having a regulatory region comprising a TGF-ß response element operatively linked to a promoter and having a structural region encoding an indicator molecule. Following TGF-ß induced activation of the TGF-ß response element, transcription results in the expression of an indicator molecule, the amount of which allows for the measurement of the amount of TGF-ß responsible for the induced activation.

In particular, in one embodiment of the invention contemplates a method for quantifying the amount of TGF-ß in a liquid sample, which method comprises:

(a) incubating the liquid sample together with eucaryotic cells that contain a TGF-ß responsive expression vector having a gene encoding luciferase for a predetermined time period sufficient for the eucaryotic cells to express a detectable amount of the luciferase;

(b) measuring the amount of the luciferase expressed

during the time period; and

(c) determining the amount of TGF-ß present in the sample by comparing the measured amount of the luciferase against a reference curve.

The invention further contemplates that the reference curve represents a quantitative relationship derived from a series of measured amounts of luciferase produced from a series of known concentrations of TGF-ß.

Another embodiment of the invention contemplates a method for quantifying the amount of transforming growth factor-ß (TGF-ß) in a liquid sample comprising:

(a) providing, in eucaryotic cells capable of expressing an indicator molecule, a plasmid comprising, in the direction of transcription, a regulatory region that includes at least one TGF-ß inducible response element that is operatively linked to a promoter, and a structural region downstream of the promoter, where the response element is capable of inducing dose-dependent indicator molecule activity and where the structural region codes for the indicator molecule;

(b) incubating the liquid sample with the eucaryotic cells for a predetermined time period sufficient for the eucaryotic cells to express a detectable amount of the indicator molecule;

(c) measuring the amount of the indicator molecule expressed during the time period; and

(d) comparing the measured amount of the indicator molecule produced in step (c) with the amount of indicator molecule produced in a control assay performed according to steps (a) through (c) by treating the liquid sample with an anti-TGF-ß antibody to obtain a net measured amount of the indicator molecule induced by TGF-ß.

Contemplated for use with the methods of this invention are plasmids having identifying characteristics of plasmids on deposit with ATCC having the ATCC Accession Numbers 75627, 75628 and 75629. Also contemplated are stably transformed

eucaryotic cells that contain the TGF-ß response element having the nucleotide sequence in SEQ ID NO 11 where the cells correspond to cells on deposit with ATCC having the ATCC Accession Number CRL 11508.

The invention describes plasmids for use in the methods that comprise a nucleotide sequence corresponding to nucleotide sequences listed in SEQ ID NOs 1-10. TGF-ß inducible response elements that comprise a nucleotide sequence corresponding to nucleotide sequences listed in SEQ ID NOs 11-17 are also described. Contemplated promoter nucleotide sequences are listed in SEQ ID NOs 18 and 19.

A further embodiment of the methods of the invention are eucaryotic cells that are stably transformed cells containing a plasmid having a gene encoding a selectable marker for the selection of said stably transformed cells. The invention describes such plasmids having nucleotide sequences listed in SEQ ID NOs 1-6. The invention further describes a stably transformed eucaryotic cell on deposit with ATCC having ATCC Accession Number CRL 11508 containing the TGF-ß response element having the nucleotide sequence in SEQ ID NO 11.

An additional embodiment are eucaryotic cells that are transiently transformed cells with plasmids corresponding to the nucleotide sequences listed in SEQ ID NOs 7-10.

The invention describes quantifying the amount of TGF-ß in a body fluid, in culture medium, and in a tissue extract. A further preferred embodiment is the determination of the amount of a specific isoform of TGF-ß, specifically TGF-ß1, TGF-ß2 or TGF-ß3, in a liquid sample.

In a preferred embodiment, this invention describes the use of mammalian cells. Preferred mammalian cells include mink lung epithelial cells, HeLa cells, Chinese hamster ovary cells, Hep3B cells, GM7373 cells, and NIH 3T3 cells.

A preferred indicator molecule also described for use with the methods of this invention is a chemiluminescent molecule, preferably luciferase.

The invention describes a composition of a plasmid vector in capable of causing expression of an indicator molecule in a eucaryotic cell, where the plasmid contains nucleotide sequences comprising a regulatory region that includes at least one TGF-ß inducible response element operatively linked to a promoter, a structural region downstream of said promoter and coding for said indicator molecule, and a gene encoding a selectable marker for the selection of a stably transformed cell, where the response element is capable of inducing dose-dependent luciferase activity.

In preferred embodiments, plasmids with selectable marker genes have the nucleotide sequences corresponding to SEQ ID NOs 1-6. Preferred TGF-ß inducible response elements for use in the expression vectors of this invention have the nucleotide sequences corresponding to SEQ ID NOs 11-17.

A further preferred embodiment of the expression vectors of this invention is the use of the neomycin gene for selecting stable transformants, the nucleotide sequence of which is listed in SEQ ID NO 20.

The invention further describes plasmids lacking a selectable marker gene having the identifying characteristics of plasmid ATCC Accession Numbers 75627, 75628, 75629, corresponding to SEQ ID NOs 8-10, respectively.

The invention describes a eucaryotic cell containing a plasmid having a nucleotide sequence listed in SEQ ID NOs 1-10.

Kits useful in assaying the amount of TGF-ß in a liquid sample comprising (a) packaging material; (b) eucaryotic cells capable of expressing an indicator molecule and containing a plasmid of this invention and an aliquot of TGF-ß, where the latter is used for generating a reference curve.

Other embodiments will be apparent to one skilled in the art.

Brief Description of the Drawings

Figure 1 shows the structure and construction of the

p800neoLuc expression vector. p800Luc was digested with AccI and blunt-ended. pMAMneo was then digested with Sal I and Eco RI, blunt-ended, and the fragment containing the neomycin-resistance gene (neo^r) was ligated to the linearized p800Luc to form p800neoLuc. Clones were analyzed via restriction enzyme mapping and one clone with the proper insert was selected. (MCS, multiple cloning site; PA1, 2, 3, polyadenylation regions 1, 2, and 3). The details of the construction are described in Example 1A.

Figure 2A, having an inset (Figure 2B), shows the dose-dependent induction of the plasminogen activator inhibitor-1/luciferase (PAI/L) construct in p800neoLuc expression vector in stably transformed MLE cells by TGF-ß1, TGF-ß2, and TGF-ß3. The TGF-ß assay was performed as described in Example 3 with DMEM-BSA containing the indicated concentrations in picomoles (pM) of recombinant (r) TGF-ß1 (closed squares), TGF-ß2 (closed circles), or TGF-ß3 (closed triangles) on the X-axis. The amount of expressed luciferase detected by a luminometer is plotted on the Y-axis and is expressed in relative light units (RLU). The results shown in Figures 2A, 2B and 2C are described in Example 3B. Figure 2B shows the treatment of p800neoLuc-transformed MLE cells with all three TGF-ß isoforms in a TGF-ß assay that resulted in a linear dose-response over the range of 0 to 4 pM of TGF-ß. In Figure 2C, the TGF-ß assay was performed with 8 pM rTGF-ß1, TGF-ß2 or TGF-ß3 in DMEM-BSA in the presence (cross-hatched bars) or absence (open bars) of 100 µg/ml of anti-TGF-ß, TGF-ß2 and TGF-ß3 monoclonal antibody. Baseline induction is indicated by medium alone (filled bars).

Figures 3A, 3B, 3C and 3D show the effects of medium, cell density and incubation time on sensitivity of the TGF-ß assay as described in Example 3B with the amount of TGF-ß1 plotted on the X-axis in pM against the measured RLU on the Y-axis. In Figure 3A, the assay was performed with increasing rTGF-ß1 concentrations in DMEM (closed squares), alpha-MEM (closed circles), CMEM (closed triangles: Eagles MEM supplemented with

non-essential amino acids) or RPMI-1640 (closed diamonds: Bio-Whittaker). All media contained 0.1% BSA. In Figure 3B, increasing concentrations of rTGF-ß1 in DMEM, 0.1% BSA were measured using 3.2×10^4 (closed squares), 1.6×10^4 (closed circles), or 0.8×10^4 (closed triangles) clone 32 (C32) of mink lung epithelial cells/well (MLE cells) after a three hour attachment period. Samples were incubated with the cells for 14 hours prior to assaying for luciferase activity. In Figures 3C and 3D (an inset in Figure 3C), 1.6×10^4 C32 cells were allowed to attach for 3 hours prior to addition of the indicated concentrations of rTGF-ß1. The samples were incubated for 6 (closed squares), 14 (closed circles), or 22 (closed triangles) hours prior to assaying for luciferase activity. The results are described in Example 3B.

Figures 4A and 4B show the effects of growth factors on the TGF-ß assay and MLEC assay while Figure 4C shows the effects caused by serum. For all figures, either the growth factors or TGF-ß are plotted on the X-axis against the RLU on the Y-axis. In Figure 4A, the TGF-ß assays were performed with DMEM-BSA containing the indicated concentrations of rTGF-ß1 (closed squares), recombinant human bFGF (closed circles), recombinant IL-1alpha (closed triangles), recombinant PDGF-BB (closed diamonds), or EGF (open squares). In Figure 4B, TGF-ß assays were performed with DMEM-BSA containing 1 pM rTGF-ß1 (closed squares) and the indicated concentrations of recombinant human bFGF (closed circles), recombinant IL-1alpha (closed triangles), recombinant PDGF (closed triangles), or EGF (open squares). The assays and results are described in Example 3C. In Figure 4C, TGF-ß assays were performed with DMEM-BSA containing the indicated concentrations of rTGF-ß1 alone (closed squares) or with 0.5% (closed circles), 1% (closed triangles), or 2% (closed diamonds) calf serum. The assays and results are described in Example 3D.

Figure 5 shows the comparison of CMs assayed by the TGF-ß (shown as the PAI/L assay) and MLEC assays. DMEM BSA (closed

squares), COS (X-marked lines), BSM (closed triangles) or BAE (closed circles) cell conditioned medium (CM) with the indicated concentrations of rTGF-ß1 were assayed by PAI/L (TGF-ß) assay (broken line) as measured by RLU on the right-hand Y-axis and MLEC (unbroken line) assay as measured by tritiated thymidine (^{3}H-thymidine) incorporation percent of controls described in Example 3E. The data points were normalized to DMEM-BSA.

Figure 6 shows the effects of growth factors on DNA synthesis as measured by ^{3}H-thymidine incorporation percent of control. In the graph, DMEM-BSA containing rTGF-ß1 (closed squares), TGF-ß2 (closed circles), TGF-ß3 (closed triangles), recombinant human bFGF (closed diamonds), recombinant IL-lalpha (open squares), EGF (open circles), or recombinant PDGF-BB (open triangles) were separately assayed using the MLEC assay as described Example 3C.

Detailed Description of the Invention

A. Definitions

Recombinant DNA (rDNA) Molecule: A DNA molecule produced by operatively linking two DNA segments. Thus, a recombinant DNA molecule is a hybrid DNA molecule comprising at least two nucleotide sequences not normally found together in nature. rDNA's not having a common biological origin, i.e., evolutionarily different, are said to be "heterologous".

Vector: A rDNA molecule capable of autonomous replication in a cell and to which a DNA segment, e.g., gene or polynucleotide, can be operatively linked so as to bring about replication of the attached segment. Vectors capable of directing the expression of genes encoding for one or more polypeptides are referred to herein as "expression vectors".

Upstream: In the direction opposite to the direction of DNA transcription, and therefore going from 5' to 3' on the non-coding strand, or 3' to 5' on the mRNA.

Downstream: Further along a DNA sequence in the direction

of sequence transcription or read out, that is traveling in a 3'- to 5'-direction along the non-coding strand of the DNA or 5'- to 3'-direction along the RNA transcript.

Reading Frame: Particular sequence of contiguous nucleotide triplets (codons) employed in translation that define the structural protein encoding-portion of a gene, or structural gene. The reading frame depends on the location of the translation initiation codon.

Response Element: Also referred to as an enhancer element, is a short DNA sequence that occurs further upstream than the upstream promoter element. Response elements contain specific nucleotide sequences recognized by transcription factors that are DNA-binding proteins.

Promoter: A region on a DNA molecule, generally from 100 to 200 base pairs longs, upstream from the coding sequence; an area to which the RNA polymerase initially binds prior to the initiation of trancription. The nucleotide sequence of the promoter, or at least part of it, determines the nature of the polymerase that associates with it. Certain consensus sequences, CAT and TATA boxes, with the promoter region are important for binding of RNA polymerase.

Regulatory Region: A DNA control module upstream from the coding sequence containing an upstream promoter element and response elements, the latter of which is also referred to as enhancer elements.

Growth Factor: A small protein that binds to a receptor for controlling cell proliferation.

Receptor: A molecule, such as a protein, glycoprotein and the like, that can specifically (non-randomly) bind to another molecule. Receptors of one type are plasma membrane proteins that bind specific molecules including growth factors, hormones, or neurotransmitters, resulting in the transmission of a signal to the cell's interior causing the cell to respond in a specific manner.

Sense Strand: A nucleotide sequence referred to as a

sense strand of a double-stranded deoxyribonucleic acid sequence is the nucleotide sequence that when read in the 5' to 3' direction by the genetic code defines an amino acid sequence of interest. Alternatively, sense strand is referred to as a coding strand.

B. Transforming Growth Factor-ß (TGF-ß)

Transforming growth factor-ß, hereinafter referred to as TGF-ß, is a growth inhibitor that exhibits a diversity of biological activities in addition to its effects on cellular proliferation. TGF-ß belongs to a large family of related molecules with a wide range of regulatory activities as described in the Background. For review, see Barnard et al., Biochim. Biophys. Acta., 1032:79-87 (1990), the disclosure of which is hereby incorporated by reference.

As previously discussed, TGF-ß is produced and secreted from cells in three distinct molecular isoforms of TGF-ß, the genes of which are located on different chromosomes, have been identified in mammals and are designated TGF-ß1, TGF-ß2 and TGF-ß3. Derynck et al., Nature, 316:701-705 (1985); Hanks et al., Proc. Natl. Acad. Sci., USA, 85:71-72 (1988); and Madisen et al., DNA, 7:1-8 (1988). Each of the isoforms are synthesized as high molecular weight latent or inactive precursor polypeptides that are then processed to 12.5 kD monomers that then dimerize to form biologically active, also referred to as mature, TGF-ß.

The activation process must occur to allow binding of the dimerized TGF-ß to the high affinity TGF-ß receptors expressed on the surfaces of all normal cells and most all neoplastic cells. Tucker et al., Proc. Natl. Acad. Sci., USA, 81:6757-6761 (1984); Frolik et al., J. Biol. Chem., 259:10995-11000 (1984); Pircher et al., Biochem. Biophys. Res. Commun., 136:30-37 (1986).

TGF-ß has been shown to induce the increase secretion of the inhibitor, plasminogen activator inhibitor-1 (PAI-1) (Laiho

et al., J. Biol. Chem., 262:17467-17474 (1987)). PAI-1 is the primary inhibitor of both tissue-type plasminogen activator (t-PA) and urokinase-type plasminogen activator (u-PA), and as such is a potent anti-fibrinolytic molecule. As a consequence of PAI-1 induction by TGF-ß, the activity of plasminogen activator (PA) is decreased. The resulting cascade of activation of plasminogen to plasmin is thereby inhibited resulting in the subsequent degradation of fibrin.

While PAI-1 synthesis by TGF-ß has been shown to occur primarily at the level of transcription following the TGF-ß receptor-ligand interaction, the mechanism of activation of the PAI-1 promoter resulting in the transcription of the PAI-1 gene is less well understood. Studies of PAI-1 gene transcription have shown that the signal transduction mechanisms are independent of de novo protein synthesis as determined by the lack of inhibition by cycloheximide and rapid onset of induction as described by Sawdey et al., J. Biol. Chem., 264:10396-10401 (1989), the disclosure of which is hereby incorporated by reference. The TGF-ß-induced enhancement of promoter activity for the alpha$_2$ collagen gene has been shown to be mediated by a binding site for nuclear factor I as described by Sporn et al., J. Cell Biol., 105:1039-1045 (1987).

As shown in Example 4, the PAI-1 promoter contains AP-1-like nucleotide sequences which is bound by the AP-1 heterodimeric transcription factor complex of Fos and Jun protein subunits. Although AP-1-like DNA enhancer sites are present in PAI-1, as shown in Example 4, activation of these sites by the AP-1 heterodimeric complex was independent of the TGF-ß-mediated induction of PAI-1 synthesis.

Although the exact transcriptional mechanism of PAI-1 promoter activation following TGF-ß receptor-ligand interaction is not known as well as the identification of the responsible TGF-ß-related transcription factor, the activation of a TGF-ß response element of this invention following TGF-ß occupancy of the TGF-ß receptor will be referred to as TGF-ß-induced

activation. Since the TGF-ß response element is activated by TGF-ß resulting in the induction of indicator protein expression, the TGF-ß response element is also referred to as a TGF-ß inducible response element

C. TGF-ß Response Elements

The present invention is based on the discovery that when eucaryotic cells, transformed with a TGF-ß-responsive expression vector of this invention, were exposed to liquid samples of TGF-ß, the resulting expression of an indicator molecule was dose-dependent in relationship to the amount of TGF-ß present in the sample. Thus, the present invention provides for a method to quantify the amount of TGF-ß in an liquid sample by measuring the amount of indicator molecules expressed.

The induced expression of the indicator molecules was the result of activation of TGF-ß response elements present in the regulatory region of the TGF-ß responsive expression vectors, the latter of which are described in Section D.

In practicing this invention, the regulation of transcription in the TGF-ß responsive expression vector-transformed eucaryotic cells is dependent TGF-ß. As described above, the TGF-ß occupation of the TGF-ß receptor expressed on the surface of cells results in the activation of a TGF-ß-related transcription factor. In general, transcription factors are site-specific DNA-binding proteins. Typically, usually positioned 5' to a structural gene is a region of nucleotide sequences that are responsible for controlling transcription. This region has been coined the "control module".

The control module comprises two categories of regulatory sequences, the promoter element and the enhancer elements. The promoter is referred to as an upstream promoter as it lies upstream of the structural genes. Promoter elements are usually 100 to 200 base pairs long and the segment of DNA is

relatively close to the site of initiation of transcription. A particular sequence recognized by one of several transcription factors that are known to bind to the promoter region is the TATA box, a region that is rich in A-T base pairs.

The enhancer regions are also referred to as response regions or response elements. Thus the term "TGF-ß response element" can also be designated "TGF-ß enhancer", "TGF-ß enhancer region", or "TGF-ß response region", and the like. The enhancer region is hereinafter referred to as a response element. They are short DNA segments that occur further upstream from the initiator site than the upstream promoter element. Response elements contain specific sequences that are recognized by transcription factors. The response elements are often a few 1000 base pairs 5' to the promoter but may even be 20,000 base pairs or more distant.

The binding of a transcription factor to either a nucleotide sequence comprising a response element or promoter resembles an "on switch". In the context of the present invention, the binding of the TGF-ß-related transcription factor results in the dose-dependent activation of the promoter resulting in the transcription of a structural region gene from DNA into RNA. In most cases, the resulting RNA molecule serves as a template for synthesis of a specific molecule, such as the indicator molecule of this invention.

Thus, "activation" of a TGF-ß response element refers to a process whereby the functional state of the TGF-ß response element is altered. The result of the TGF-ß activation of the TGF-ß response element is an increase in the transcriptional efficiency of the structural gene driven from the promoter.

A further embodiment of a TGF-ß response element is that it is inducible. The term "inducible" refers to a an enhancement of a particular function. In this invention, the functional activity of a TGF-ß response element is increased or induced following activation by the TGF-ß-related transcription factor. Thus, the TGF-ß response element is also referred to

as a TGF-ß inducible response element.

The result of TGF-ß response element activation is the coordinate transcription and translation of the structural region containing a gene encoding an indicator protein of this invention as described in Section D. The resulting expression of an indicator molecule is dose-dependent in relationship to the amount of TGF-ß present in the sample.

The term "dose-dependent" refers to the functional relationship between the amount of TGF-ß activating the TGF-ß response element and the resulting expression of the indicator molecule. Thus, the functional relationship between TGF-ß activation and expression of an indicator molecule can be referred to as a linear relationship. Because of the dose-dependent expression of an indicator molecule, such as luciferase, in response to TGF-ß exposure, the amount of TGF-ß responsible for the activation of the expression can be readily determined using the methods of this invention.

Thus, based on the teachings herein, a TGF-ß response element nucleotide sequence is characterized by its ability to be responsive to TGF-ß-induced activation. Such a TGF-ß response element is useful herein as a component in the expression vectors of this invention to provide for the ability to quantify the amount of TGF-ß responsible for the transcriptional activation. Thus, a TGF-ß response element of this invention comprises any nucleotide sequence that is activated by TGF-ß, the process of which is as described in Section B.

In the context of this invention, the term nucleotide sequence refers to a plurality of joined nucleotide units formed from naturally- or non-naturally occurring bases and cyclofuranosyl groups joined by phosphodiester bonds. Thus, the nucleotide sequence includes the use of nucleotide analogs.

One embodiment of a TGF-ß response element of this invention is an isolated double-stranded deoxyribonucleic acid molecule comprising a sequence of nucleotide bases that defines

a TGF-ß response element. However, neither is it necessary that the obtained TGF-ß be a naturally occurring sequence present in the other genes nor that the TGF-ß response element be limited to deoxyribonucleotides. The TGF-ß response element may be found in DNA or RNA, in regulatory sequences, exons, or introns.

Preferred TGF-ß response elements are derived from selected regions of the promoter regions of the plasminogen activator inhibitor type 1 gene, hereinafter referred to as PAI-1, as described by Loskutoff et al., Biochem., 26:3763-3768 (1987), the disclosure of which is hereby incorporated by reference. Loskutoff et al. describes a cosmid containing the entire PAI-1 gene. In a related study, the glucocorticoid regulation of the PAI-1 promoter was described by van Zonneveld et al., Proc. Natl. Acad. Sci., 85:5525-5529 (1988), the disclosure of which is hereby incorporated by reference. The sequence of the PAI-1 promoter corresponding to nucleotide positions -800 and extending through the TATA box and initiation site and ending at nucleotide position +200, the latter of which corresponds to the PAI-1 encoded protein at the ninth amino acid residue, in available in the GenBank™/EMBL Data Bank with Accession Number J03836.

Moreover, Bosma et al., J. Biol. Chem., 263:9129-9141 (1986), have described the entire 15,867 bp PAI-1 gene sequence including significant stretches of DNA that extend into its 5'- and 3'-flanking DNA regions, the nucleotide sequence of which is available in the GenBank™/EMBL Data Bank with Accession Number J03764.

The PAI-1 promoter-derived TGF-ß response elements for use in this invention are identified by the nucleotide positions corresponding to the region in the PAI-1 promoter as listed in the GenBank™/EMBL Data Bank Accession Number J03836.

Exemplary TGF-ß response elements derived from the PAI-1 promoter have the nucleotide sequences listed in the Sequence Listing in SEQ ID NOs 11-17. The nucleotide sequences are

listed showing only the sense strand in the 5' to 3' direction of a double-stranded isolated TGF-ß response element nucleotide sequence. The PAI-1-derived TGF-ß response elements corresponding to SEQ ID NOs 11-17 have the respective designations with the nucleotide regions corresponding to the PAI-1 promoter indicated in parentheses: 1) SEQ ID NO 11 = 1500 (-1481 to -40); 2) SEQ ID NO 12 = 800 (-800 up to -40); 3) SEQ ID NO 13 = 800/636 (-800 up to -636); 4) SEQ ID NO 14 = 56 (-56 to -41); 5) SEQ ID NO 15 = 674 (-674 to -650); 6) SEQ ID NO 16 = 743 (-743 to -708); and 7) SEQ ID NO 17 = 732 (-732 to -708).

In one embodiment, a TGF-ß response element useful for practicing the present invention may be derived from any promoter nucleotide sequence. In a further embodiment, a TGF-ß response element may be designed to contain preselected nucleotide bases. In other words, a subject TGF-ß response element need not be identical to the nucleotide sequence of the PAI-1-derived TGF-ß response elements described herein, so long as the nucleotide sequence is activatable by TGF-ß.

A TGF-ß response element of this invention thus may contain a variety of nucleotide units of any length, typically from about 5 to about 2000 nucleotides in length. More preferably, a TGF-ß response element comprises nucleotide units from about 15 to about 1500 nucleotides in length.

A preferred embodiment is a TGF-ß response element having nucleotide sequences that is greater than 50 base pairs in length. Exemplary long TGF-ß response elements derived from PAI-1 are listing in the Sequence Listing in SEQ ID NOs 11-13.

A preferred embodiment is a TGF-ß response element having nucleotide sequences that is less than 50 base pairs in length. Exemplary short TGF-ß response elements derived from PAI-1 are listing in the Sequence Listing in SEQ ID NOs 14-17.

In one embodiment, the invention contemplates the presence of at least one TGF-ß response element present in the regulatory region of the expression vectors as described in

Section D. Thus, one or more stretches of a nucleotide sequence comprising a TGF-ß response element may be present within a regulatory region. If more than one TGF-ß response element is present, they are not required to be identical. In other words, TGF-ß response elements having different nucleotide sequences as well as different lengths can be combined in a regulatory region of an expression vector of this invention.

TGF-ß response elements can be derived or produced from the PAI-1 promoter by truncation or expansion of the native or wild-type PAI-1 promoter nucleotide sequence or as a variant of the native PAI-1 promoter by site-directed substitution of a preselected nucleotide base or bases.

Also contemplated in this context are regulatory regions containing multiple TGF-ß response elements that can be either longer, shorter, tandemly arranged, reversed in orientation, and permutations thereof. The design and construction of such arrangements are well known to one of ordinary skill in the art of oligonucleotide design and synthesis and are described by Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory, pp 390-401 (1982).

It is also contemplated that nucleotide base modifications can be made resulting in nucleotide analogs to provide certain advantages to the TGF-ß response elements of this invention.

A nucleotide analog refers to moieties that function similarly to nucleotide sequences in a TGF-ß response element of this invention but which have non-naturally occurring portions. Thus, nucleotide analogs can have altered sugar moieties or inter-sugar linkages. Exemplary are the phosphorothioate and other sulfur-containing species, analogs having altered base units, or other modifications consistent with the spirit of this invention.

Preferred modifications include, but are not limited to, the ethyl or methyl phosphonate modifications disclosed in the U.S. Patent No., 4,469,863 and the phosphorothioate modified

deoxyribonucleotides described by LaPlanche et al., Nucl. Acids Res., 14:9081 (1986) and Stec et al., J. Am. Chem. Soc., 106:6077 (1984), the disclosures of which are hereby incorporated by reference. These modifications provide resistance to nucleolytic degradation. Preferred modifications are the modifications of the 3'-terminus using phosphothionate (PS) sulfurization modification described by Stein et al., Nucl. Acids Res., 16:3209 (1988).

TGF-ß response elements comprising nucleotide sequences can be obtained by a variety of procedures well known in the art, including de novo chemical synthesis of complementary oligonucleotides and derivation of nucleic acid fragments from native nucleic acid sequences existing as genes, or parts of genes, in a genome, plasmid, or other vector, such as by restriction endonuclease digestion of larger nucleic acid fragments and strand separation or by enzymatic synthesis using a nucleic acid template.

De novo chemical synthesis of oligonucleotides can be carried out, for example, by the phosphotriester method described by Matteucci et al., J. Am. Chem. Soc., 103:3185 (1981), or as described in U.S. Patent No. 4,356,270, the disclosures of which are hereby incorporated by reference. A particularly preferred method is the phosphoramide method using commercial automated synthesizers, such as the ABI automated synthesizer by Applied Biosystems. Inc., (Foster City, CA). Oligonucleotides can be purified after synthesis using published procedures as described by Miller et al., J. Biol. Chem., 255:9659 (1980). Thereafter, complementary oligonucleotides are hybridized to form double-stranded DNA segments that are TGF-ß response elements. Particularly preferred chemically-synthesized oligonucleotides are described in Example 1C and the sense strands of which are listed in SEQ ID NOs 14-17, as described above.

Derivation of a TGF-ß response element from nucleic acids involves the cloning of a nucleic acid into an appropriate host

by means of a cloning vector, replication of the vector and therefore multiplication of the amount of the cloned nucleic acid followed by isolation of subfragments of the cloned nucleic acids. For a description of subcloning nucleic acid fragments, see Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory, pp 390-401 (1982); and see U.S. Patent Nos 4,416,988 and 4,403,036.

In one embodiment, TGF-ß response elements are obtained by restriction digestion of cloned vectors containing the PAI-1 promoter as described in Example 1A and 1C. Particularly preferred nucleotide sequences containing TGF-ß response elements as well as the minimal promoter sequence obtained in this manner include nucleotide sequences corresponding to the nucleotide positions in the PAI-1 promoter sequence from -1481 to +76, specifically a Kpn I/Eco RI digest and -800 to +76, specifically a Hind III/Eco RI digest.

In an additional embodiment, in the practice of this invention, it is not necessary that the TGF-ß response element nucleotide sequence be known in order to obtain a TGF-ß response element capable of being activated by TGF-ß. To that end, contemplated for use in this invention are TGF-ß response elements obtained from promoter regions of other genes that can be determined to contain TGF-ß response elements using the methods of this invention.

D. TGF-ß Responsive Plasmid Expression Vectors

The present invention contemplates TGF-ß responsive plasmid expression vectors in substantially pure form capable of causing expression of an indicator molecule in a eucaryotic cell. The term "TGF-ß responsive" identifies an expression vector of this invention that by its composition contains TGF-ß response elements that are activated by TGF-ß mediated through a TGF-ß response element specific transcription factor as described in Section C. Vectors capable of directing the expression of genes to which they are operatively linked are

referred to herein as "expression vectors".

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting between different genetic environments another nucleic acid to which it has been operatively linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked.

A TGF-ß expression vector of this invention is a circular double-stranded plasmid that contains at least the following elements: 1) a regulatory region having at least one TGF-ß response element as defined in Section C, where the regulatory region is operatively linked to a promoter; and 2) a structural region downstream of the promoter that contains a gene coding for an indicator molecule of this invention.

In a separate embodiment, a TGF-ß expression vector also contains a gene, the expression of which confers a selective advantage, such as a drug resistance, to the eucaryotic host cell when introduced or transformed into those cells. A typical eucaryotic drug resistance genes confers resistance to neomycin, also referred to as G418 or Geneticin.

The choice of vector to which the regulatory region, promoter, and structural region of the present invention is operatively linked depends directly, as is well known in the art, on the functional properties desired, e.g., replication or protein expression, and the host cell to be transformed, these being limitations inherit in the art of constructing recombinant DNA molecules.

In preferred embodiments, the vector utilized includes procaryotic sequences that facilitate the propagation of the vector in bacteria, i.e., a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extra-chromosomally when introduced into a bacterial host cell. Such replicons are well known in

the art. In addition, the TGF-ß expression vector of this invention includes one or more transcription units that are expressed only in eucaryotic cells.

The eucaryotic transcription unit consists of noncoding sequences and sequences encoding selectable markers. The expression vectors of this invention also contain distinct sequence elements that are required for accurate and efficient polyadenylation, referred to as PA1, 2 and 3 and as shown in Figure 1. In addition, splicing signals for generating mature mRNA are included in the vector. The eucaryotic TGF-ß responsive expression vectors contain viral replicons, the presence of which provides for the increase in the level of expression of cloned genes. A preferred replication sequence is provided by the simian virus 40 or SV40 papovavirus.

Operatively linking refers to the covalent joining of nucleotide sequences, preferably by conventional phosphodiester bonds, into one strand of DNA, whether in single- or double-stranded form. Moreover, the joining of nucleotide sequences results in the joining of functional elements such as response elements in regulatory regions with promoters and downstream structural regions as described herein.

A preferred eucaryotic expression vector of this invention as prepared in Example 1 contains a regulatory region having TGF-ß response elements derived from the 5' promoter end of the human plasminogen activator inhibitor type 1 (PAI-1) gene operatively linked to PAI-1 minimal promoter and a downstream structural region containing a gene coding for an indicator polypeptide, preferably luciferase.

Exemplary TGF-ß responsive expression vectors include the following expression vectors, the designations of which are indicated along with the corresponding SEQ ID NO in which the sense strand of the expression vector is listed where the first nucleotide of the double-stranded circular vector is the middle "T" nucleotide present in the Eco RI restriction site as described in Example 1: 1) p800neoLuc (SEQ ID NO 1); 2)

p800/636neoLuc (SEQ ID NO 2); 3) p56neoLuc (SEQ ID NO 3); 4) p674neoLuc (SEQ ID NO 4); 5) p743neoLuc (SEQ ID NO 5); 6) p732neoLuc (SEQ ID NO 6); 7) p56Luc (SEQ ID NO 7); 8) p674Luc (SEQ ID NO 8); 9) p743Luc (SEQ ID NO 9); and 10) p732Luc (SEQ ID NO 10).

The exemplary TGF-ß expression vectors of this invention are derived from the starting cloning expression vector, designated p19Luc, as described in Example 1. The nucleotide sequence of the sense strand of an Eco RI-linearized p19LUC vector is listed in the Sequence Listing as SEQ ID NO 21.

A further embodiment of this invention is the preparation of TGF-ß responsive expression vectors having altered arrangements of and selected types of TGF-ß response elements in the regulatory region. To that end, p19Luc and the p19Luc-derived p39Luc expression cloning vectors, both of which is described in Example 1, are vectors that allow for the construction of TGF-ß responsive vectors having any selected regulatory region operatively ligated to a selected promoter. Therefore, any regulatory region of any length containing one or more TGF-ß response elements can be paired with any promoter, a non-TGF-ß responsive PAI-1 or heterologous HBV promoter as used herein but not limited to that, to prepare TGF-ß responsive expression vectors that provide for the quantitation of inducing TGF-ß.

In a related embodiment, in addition to the construction methods detailed herein, other methods of preparing p19Luc-derived expression vectors having TGF-ß response elements and promoters are familiar to one of ordinary skill in the art of vector construction and are described by Ausebel, et al., In Current Protocols in Molecular Biology, Wiley and Sons, New York (1993) and by Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, 1989.

1. Plasmid Vectors for Stable Transformations

In practicing one aspect of this invention, a

preferred embodiment is a TGF-ß responsive expression vector having a gene for encoding a selectable marker providing for stably transformed cells. Stably transformed cells confer the ability to utilize a reproducible source for practicing the methods of this invention over a course of time. A preferred selectable marker gene is the gene conferring neomycin-resistance. Such a gene for encoding the selectable marker was derived from an expression vector, designated pMAMneo, as described in Example 1. The nucleotide sequence of the neomycin-resistance conferring gene is listed in SEQ ID NO 20.

In one embodiment, a TGF-ß responsive expression vector contains a first nucleotide sequence comprising a regulatory region that includes at least one TGF-ß inducible response element operatively linked to a promoter, a second nucleotide sequence comprising a structural region downstream of the promoter and coding for an indicator molecule, and a third nucleotide sequence comprising a gene encoding a selectable marker for the selection of a stably transformed cell, where the response element is capable of inducing dose-dependent luciferase activity and the structural region codes for luciferase.

Preferred expression vectors containing the neomycin-resistance conferring gene include the following designations followed in parenthesis by the corresponding SEQ ID NO in which the sense strand of each Eco RI-linearized vector is listed according to the convention adopted in this invention for listing vector sequences: 1) p800neoLuc (SEQ ID NO 1); 2) p800/636neoLuc (SEQ ID NO 2); 3) p56neoLuc (SEQ ID NO 3); 4) p674neoLuc (SEQ ID NO 4); 5) p743neoLuc (SEQ ID NO 5); 6) p732neoLuc (SEQ ID NO 6).

In a further embodiment, the plasmid expression vectors of this invention contain TGF-ß inducible response elements that correspond to a nucleotide sequence listed in SEQ ID NOs 11-17 as described in Section C.

Preferred promoters for use in the TGF-ß expression

vectors of this invention for stably transforming cells as well as for transient transformation are the PAI-1 minimal promoter sequence and the hepatitis B virus minimal promoter sequence, the sense sequences of which are respectively listed in SEQ ID NOs 18 and 19. Contemplated for use in this invention are promoters that are not responsive to TGF-ß. The selection of alternative promoters is within the scope of one having ordinary skill in the art.

This invention contemplates additional TGF-ß expression vectors for stably transforming cells that can be designed to have regulatory regions that contain alternative TGF-ß response elements and promoters.

a. Regulatory Region

The regulatory region of a TGF-ß expression vector of this invention contains at least one TGF-ß response element as described herein and in Section C of this invention. As contemplated for use in this invention, the regulatory region of a TGF-ß expression vector can range in length from 5 to 2000 base pairs, preferably 15 to 1500 base pairs, and can contain more than one TGF-ß response element in any orientation and arrangement. Thus, if two or more TGF-ß response elements are present in a regulatory region, they may be contiguous with one another or separated by an intervening nucleotide sequence. The design and construction of such arrangements are well known to one of ordinary skill in the art of oligonucleotide design and synthesis and are described by Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory, pp 390-401 (1982).

Preferred TGF-ß response elements present in the regulatory region of a TGF-ß expression vector are derived from the PAI-1 promoter and have the nucleotide sequences listed in the Sequence Listing in SEQ ID NOs 11-17. The PAI-1-derived TGF-ß response elements corresponding to SEQ ID NOs 11-17 have the respective designations with the nucleotide regions

corresponding to the PAI-1 promoter indicated in parentheses: 1) SEQ ID NO 11 = 1500 (-1481 to -40); 2) SEQ ID NO 12 = 800 (-800 up to -40); 3) SEQ ID NO 13 = 800/636 (-800 up to -636); 4) SEQ ID NO 14 = 56 (-56 to -41); 5) SEQ ID NO 15 = 674 (-674 to -650); 6) SEQ ID NO 16 = 743 (-743 to -708); and 7) SEQ ID NO 17 = 732 (-732 to -708).

b. Structural Region

A plasmid vector of the present invention contain a structural region having a nucleotide sequence that encodes an indicator molecule. The structural region is operatively linked to the regulatory region such that the inducible promoter of the regulatory region, under the inducible control of the TGF-ß response element, controls transcription and expression of the indicator molecule. Thus, upon induction of the TGF-ß response element, the regulatory region transcribes and thereby expresses the indicator molecule resulting in a detectable event in the cell, which event can be measured by detection of the amount of the expressed indicator molecule. In other words, the response element is capable of inducing the expression of the indicator molecule by virtue of it's controlling expression of the indicator through the promoter to which the response element is operatively linked.

Typically, the structural region is "downstream" of the regulatory region in the plasmid, and positioned to be under the direct control of the regulatory region. Other configurations can be utilized so long as the induction of the TGF-ß response element results in the expression of the indicator polypeptide. Exemplary and preferred configurations are described in Examples.

The term "indicator molecule" as used in this invention refers to a molecule encoded by a reporter gene, the expression of which in the expression vectors of this invention, results in a detectable measurable protein, polypeptide, enzyme and the like. Alternative expressions for indicator molecule are

reporter molecule, reporter polypeptide, indicator protein, indicator polypeptide and the like. In preferred embodiments, the indicator molecule is a protein.

There are any of a variety of indicator polypeptides suitable for use in the present invention, and the invention need not be so limited to any particular indicator. A preferred indicator polypeptide is luciferase encoded by the firefly luciferase gene. Use of the luciferase gene for expression of luciferase has been described by Gould et al., Anal. Biochem., 7:5-13 (1988) and Brasier et al., Bio-Techniques, 7:1116-1122 (1989). A preferred structural region includes a nucleotide sequence having the sequence characteristics of the luciferase gene shown in SEQ ID NO 21.

Alternative embodiments include indicator proteins such a ß-galactosidase and chloramphenicol acetyltransferase (CAT). Use of a ß-galactosidase and CAT as reporter molecules have been respectively by Luskin et al., Neuron, 1:635-647 (1988) and Gorman et al., Mol. Cell Biol., 2:1044-1051 (1982).

Associated with the use of an indicator molecule in the quantifying TGF-ß are means for measuring the indicator molecule. A preferred method for detecting the luciferase indicator molecule is the use of a luminometer commercially available from Dynatech Laboratories Inc., Chantilly, VA as described in Example 3A and analyzed according to manufacturer's instructions. For detecting CAT activity, a simple-phase extraction assay has been developed and described by Seed et al., Gene, 67:271-277 (1988), the disclosure of which is hereby incorporated by reference. Alternative preferred methods for detecting CAT activity are described in Current Protocols in Molecular Biology, Eds, Ausebel et al., Unit 9.0, John Wiley & Sons (1993). Expression of ß-galactosidase activity is performed in activity assays performed essentially as described by Miller, Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, New York, (1972), the disclosure of which is hereby incorporated by

reference. With ß-galactosidase additional reagents are required to visualize its presence following induced expression. Such additional reagents for ß-galactosidase include o-nitrophenyl-ß-D-galactopyransoside and the like for the development of a color reaction by absorbance at wavelengths of 500 and 420.

c. Selectable Marker Gene

In preferred embodiments, the plasmid vector of the present invention includes a gene that encodes a selectable marker that is effective in a eucaryotic cell, preferably a drug resistance selection marker. A preferred drug resistance selection marker is a gene whose expression results in neomycin resistance, i.e., the neomycin phosphotransferase (neo) gene [Southern et al., J. Mol. Appl. Genet., 1:327-341 (1982)] or a gene whose expression results kanamycin resistance, i.e., the chimeric gene containing nopaline synthetase promoter, Tn5 neomycin phosphotransferase II and nopaline synthetase 3' non-translated region described by Rogers et al., Methods for Plant Molecular Biology, A. Weissbach and H. Weissbach, eds., Academic Press, Inc., San Diego, CA (1988). Other selectable markers which are utilizable in eucaryotic cells can be utilized in the present vectors and methods and therefore the invention need not be limited to any particular selectable marker. Thus, the invention contemplates the use of a nucleotide sequence which confers a eucaryotic selection means, including but not limited to genes for resistance to neomycin and kanamycin.

A preferred nucleotide sequence defining a selectable marker gene is a nucleotide sequence having the sequence characteristics of the neomycin resistance gene shown in SEQ ID NO 20.

The use of a selectable marker for eucaryotic cells provides the advantage of producing stably transformed cells, as discussed herein. Thus, one can produce a eucaryotic cell

line containing a plasmid vector of this invention for use in the present methods wherein all the cells of the culture are selected to be uniform and each contain intact plasmid vector, thereby assuring that all of the eucaryotic cell in the culture are substantially similar in responsiveness to TGF-ß, thereby increasing the reliability and sensitivity of the assay.

In addition, preferred embodiments that include a procaryotic replicon also include a gene whose expression confers a selective advantage, such as a drug resistance, to the bacterial host cell when introduced into those transformed cells. Typical bacterial drug resistance genes are those that confer resistance to ampicillin or tetracycline.

Those vectors that include a procaryotic replicon also typically include convenient restriction sites for insertion of a recombinant DNA molecule of the present invention. Typical of such vector plasmids are pUC8, pUC9, pBR322, and pBR329 available from BioRad Laboratories, (Richmond, CA) and pPL, pK and K223 available from Pharmacia, (Piscataway, NJ), and pBLUESCRIPT and pBS available from Stratagene, (La Jolla, CA). A vector of the present invention may also be a Lambda phage vector including those Lambda vectors described in Molecular Cloning: A Laboratory Manual, Second Edition, Maniatis et al., eds., Cold Spring Harbor, NY (1989).

Plasmid vectors for use in the present invention are also compatible with eukaryotic cells. Eucaryotic cell expression vectors are well known in the art and are available from several commercial sources. Typically, such vectors provide convenient restriction sites for insertion of the desired recombinant DNA molecule, and further contain promoters for expression of the encoded genes which are capable of expression in the eucaryotic cell, as discussed earlier. Typical of such vectors are pSVO and pKSV-10 (Pharmacia), and pPVV-1/PML2d (International Biotechnology, Inc.), and pTDT1 (ATCC, No. 31255).

2. Plasmid Vectors for Co-transformation and Transient Transformation

This invention contemplates the use of TGF-ß responsive expression vectors having regulatory, promoter and structural regions but lacking a gene for encoding a selectable marker. In other words, in practicing this invention, TGF-ß expression vectors for transient transformation of eucaryotic cells are contemplated. This embodiment allows for an alternative to stable transformation of cells for use practicing the methods of this invention. Transiently transformed cells produced as described in Example 2D. are useful for performing TGF-ß assays when having stably transformed cells is not required or necessitated. As described in Example 4, transiently transformed cells are useful for determining the nucleotide sequence of TGF-ß response elements as well as quantifying the amount of TGF-ß present in a heterogeneous or homogeneous liquid sample.

Preferred TGF-ß expression vectors used for transiently transforming eucaryotic cells include the following vectors shown with their designations and SEQ ID NOs in which the sense strand of the double-stranded Eco RI-linearized vectors is listed: 1) p56Luc (SEQ ID NO 7); 2) p674Luc (SEQ ID NO 8); 3) p743Luc (SEQ ID NO 9); and 4) p732Luc (SEQ ID NO 10).

The invention further describes TGF-ß responsive plasmids lacking a selectable marker gene having the identifying characteristics of plasmids that have been deposited with the American Type Culture Collection, Rockville, MD having the assigned ATCC Accession Numbers 75627, 75628, 75629, the plasmids of which respectively correspond to the Eco RI-linearized sense strand nucleotide sequences listed SEQ ID NOs 8-10.

In an additional embodiment, this invention describes the co-transformation of TGF-ß expression vectors for transient transformation in conjunction with a second expression vector from which a selectable marker is expressed. A preferred

selectable marker expressing plasmid is RSVneo as described in Example 2C. The ability to prepare stably transformed cells through the use of a vector that only confers transient transformation is accomplished with this approach. The advantage this approach provides is that further vector constructions for inserting selectable marker genes can be avoided, thereby providing stably transformed cells for use in practicing this invention when necessitated. Thus, eucaryotic cells that have been co-transformed with a transient TGF-ß expression vector and a second plasmid such as RSVneo provide for an alternative approach to create stably transformed eucaryotic cells.

Any transient TGF-ß expression vector of this invention can be used in this context. A preferred co-transformed eucaryotic cell is the cell line Hep3B that has been co-transformed with RSVneo and the p1500Luc expression vector having the TGF-ß response element in SEQ ID NO 11. This stably transformed cell line has been deposited with the American Type Culture Collection, Rockville, MD and has been assigned ATCC having ATCC Accession Number CRL 11508.

With the teachings of this invention, additional TGF-ß expression vectors for transiently transforming cells can be designed to have regulatory regions that contain alternative TGF-ß response elements and promoters. In a further embodiment, these additional vectors can be used to prepare stably transformed cells through the use of the co-transformation approach.

3. Recipient Cells for Transformations

Insofar as the invention describes plasmid vectors for use in the present invention, the invention also contemplates a eucaryotic cell containing a plasmid vector of the present invention.

A eucaryotic cell suitable for use can be any eucaryotic cell which expresses a TGF-ß receptor on its cell surface and

is capable of induction of a TGF-ß response element. There are a variety of means to identify a suitable eucaryotic cell, including, but not limited to transformation by a plasmid vector of this invention, followed by assay for expression of the indicator polypeptide upon challenge by TGF-ß.

In a preferred embodiment, this invention contemplates the use of mammalian cells. Preferred mammalian cells include mink lung epithelial cells, HeLa cells, Chinese hamster ovary cells, Hep3B cells, GM7373 cells, NIH 3T3 cells, and the like cells. These and other suitable mammalian cells are widely available. Suitable mammalian cells for use in the invention can also be obtained from the American Type Culture Collection (ATCC; Rockville, MD).

Introduction of a plasmid vector of the present invention into a eucaryotic cell can be accomplished by a variety of methods well known in the art, including, but not limited to transfection, transformation, electroporation, microinjection, liposome fusion, and the like introduction methods. Such methods are well known and are not to be considered essential to the invention. Furthermore, the introduction of the plasmid vector can be transient or stable.

A transient introduction is one where there is no selection to maintain the plasmid vector within the host eucaryotic cell through multiple rounds of cell division. Therefore, the assay is to be conducted in a short time period after introduction, and before several rounds of cell division. Stable introduction of plasmid involves the culturing of the cell under conditions that select for the maintenance of the plasmid vector, typically by the use of a gene on the plasmid that encodes a selectable marker, as described further herein.

Following the introduction of the plasmid vector, the resulting eucaryotic cell containing a plasmid vector is used in the assay methods described herein. A preferred eucaryotic cell contains a plasmid vector of this invention, which plasmid vector comprises a nucleotide sequence having a TGF-ß response

element and a gene encoding an indicator polypeptide, wherein the plasmid is capable of expression of the indicator polypeptide in response to TGF-ß induction. Particularly preferred are eucaryotic cells that contain a plasmid vector having a nucleotide sequence with the nucleotide sequence characteristics of the TGF-ß response element selected from the group consisting of the sequences shown in SEQ ID NOs 11-17. A particularly preferred eucaryotic cell contains a plasmid vector having a nucleotide sequence with the nucleotide sequence characteristics of the plasmid vector selected from the group consisting of the sequences shown in SEQ ID NOs 1-10.

A preferred eucaryotic cell described further herein is Hep3B stably transformed with the plasmid vector p1500Luc, referred to as LUCI, and having the ATCC accession No. CRL 11508.

E. Methods for Quantifying TGF-ß

The present invention describes methods for detecting the presence, and preferably quantifying the amount, of TGF-ß in a liquid sample, either containing purified TGF-ß or TGF-ß in a heterogeneous admixture, and is also referred to herein as a TGF-ß assay. The assay system provides for the quantification of TGF-ß through the expression of an indicator polypeptide which is expressed in levels proportional to the amount of TGF-ß being detected.

The assay is a highly sensitive and specific, non-radioactive assay, for detecting mature (active) TGF-ß. When compared to the sensitive and widely used proliferation-based mink lung epithelial cell (MLE cells) method for measuring TGF-ß concentration, the TGF-ß assay method of this invention is more rapid, has comparable sensitivity, and has a greater detection range. Specificity of this novel assay was also higher as evidenced by its relative insensitivity to factors such as epidermal growth factor (EGF) and basic fibroblast growth factor (bFGF) which can greatly affect other assays.

The use of a TGF-ß response element, such as the truncated PAI-1 promoter, that does not respond to other growth modulators such as platelet-derived growth factor (PDGF) found in biological samples provides an added advantage that the method of this invention can be used in conditions where other bioassays are difficult to interpret. Because of its large range and specificity, the rapid, sensitive, non-radioactive, easily performed assay method of this invention is useful in determining active TGF-ß concentrations in complex solutions.

Thus, the present invention overcomes the limitations of existing methods used to quantify the amount of TGF-ß in a liquid sample. This invention contemplates a method for quantifying the amount of TGF-ß in a sample using a system comprising a TGF-ß responsive cell containing an expression vector having a TGF-ß response element and an indicator molecule. Following TGF-ß induction, transcription results in the expression of an indicator molecule, the amount of which allows for the measurement of the amount of TGF-ß responsible for the induction.

TGF-ß receptor-bearing cells are transfected with a TGF-ß responsive expression vector of this invention, and are subsequently exposed to TGF-ß whereupon the TGF-ß receptor-bearing cells activate the TGF-ß response element in the vector which results in the concomitant expression of the indicator polypeptide. The resulting expressed indicator polypeptide is then measured in a manner depending upon the indicator polypeptide employed.

The measured indicator polypeptide resulting from activation by TGF-ß in the test liquid sample is then compared to a standardized reference curve produced using known amounts of TGF-ß.

In particular, one embodiment of the invention contemplates a method for quantifying the amount of TGF-ß in a liquid sample, which method comprises:

(a) incubating the liquid sample together with eucaryotic

cells that contain a TGF-ß responsive expression vector having a gene encoding an indicator polypeptide for a predetermined time period sufficient for the eucaryotic cells to express a detectable amount of the indicator polypeptide;

(b) measuring the amount of the indicator polypeptide expressed during the time period; and

(c) determining the amount of TGF-ß present in the sample by comparing the measured amount of the indicator polypeptide against a reference curve.

Preferably, the reference curve represents a quantitative relationship derived from a series of measured amounts of indicator polypeptide produced from a series of known concentrations of TGF-ß.

The standardized reference curve is obtained from parallel assays performed by exposing similarly transfected cells to a range, usually in serial dilution, of known (measured) amounts of one or more of the known TGF-ß isoforms. The resulting expressed indicator polypeptide is then determined by direct detection of the indicator polypeptide. A reference curve is then generated by plotting the measured amount of expressed indicator polypeptide against the known range of inducing amounts of TGF-ß. The amount of unknown TGF-ß in the test liquid sample is then determined by extrapolating the measured amount of test indicator polypeptide to the reference curve.

The use of standard curves in quantifying the amount of protein in a liquid sample in general has been described by Lowry et al., J. Biol. Chem., 193:265-275 (1951), the disclosure of which is hereby incorporated by reference. As shown in the Examples herein, the TGF-ß assay of this invention allows for the measurement of TGF-ß from the expression and subsequent detection of an indicator polypeptide from a concentration range from less than 5 picograms/ml (pg/ml) equivalent to 0.2 pM up to 10 ng/ml equivalent to 40 pM (or 0.4 nM). The dose-dependent response to TGF-ß is linear between 0.2 pM up to 100 pM depending on the assay conditions.

As described further herein, any of a variety of indicator polypeptides can be utilized in the present methods, and the invention is not to be construed as limited to any particular indicator polypeptide. However, a preferred embodiment utilizes a chemiluminescent molecule, more preferably luciferase, as the indicator polypeptide, and therefore the examples herein using luciferase are to be considered exemplary of all indicator polypeptides and of preferred embodiments. The level of expressed luciferase is easily and conveniently measured using a luminometer as described herein.

In another embodiment of the present invention, the assay method for quantifying TGF-ß in complex solutions is practiced generally as described above, but with the additional use of a neutralizing anti-TGF-ß monoclonal antibody admixed with the test liquid sample in assays run in parallel to untreated test liquid samples as described in Example 3B. These control assays are used to determine if other molecules are present in the test sample that can affect the assay through either inhibition or activation of other regions of the TGF-ß response element. For example, conditioned medium obtained from cell cultures and body fluids contain growth factors and DNA binding proteins that function as transcriptional activators or inhibitors. If a corresponding response element for an additional non-TGF-ß activator is present in the expression vector, the binding of the activator to the response element may cause enhanced or diminished expression of the indicator polypeptide. By antibody neutralization of the TGF-ß in the test sample, any residual measured indicator polypeptide can then be ascribed to non-TGF-ß activation.

The shorter TGF-ß response elements used in the expression vector systems of this invention are less likely to have non-TGF-ß response elements as shown in Examples 3E and 3F. Thus, the use of parallel antibody control assays to allow for a determination of the amount of luciferase produced from only TGF-ß activation is preferred when using expression vectors

having longer response elements or elements likely to exhibit responsiveness to transcription factors other that those induced by TGF-ß. Moreover, while the TGF-ß assay is not generally isoform specific, The assay can be TGF-ß isoform-specific by the use of the appropriate standard reference curves and parallel assays with neutralizing antibodies immunospecific to a particular TGF-ß isoform species, thereby allowing for quantification of unique TGF-ß isoforms.

Thus, in another embodiment of the invention, a method for quantifying the amount of transforming growth factor-ß (TGF-ß) in a liquid sample is contemplated, the method comprising:

(a) providing, in eucaryotic cells capable of expressing an indicator molecule, a plasmid comprising, in the direction of transcription, a regulatory region that includes at least one TGF-ß inducible response element that is operably linked to a promoter, and a structural region downstream of the promoter, where the response element is capable of inducing dose-dependent indicator molecule activity and where the structural region codes for the indicator molecule;

(b) incubating the liquid sample with the eucaryotic cells for a predetermined time period sufficient for the eucaryotic cells to express a detectable amount of the indicator molecule;

(c) measuring the amount of the indicator molecule expressed during the time period; and

(d) comparing the measured amount of the indicator molecule produced in step (c) with the amount of indicator molecule produced in a control assay performed according to steps (a) through (c) by treating the liquid sample with an anti-TGF-ß antibody to obtain a net measured amount of the indicator molecule induced by TGF-ß.

The use of a monoclonal antibody specific for TGF-ß provides particular advantages in practicing the invention. First, one can use a variety of TGF-ß response elements,

including those which exhibit responsiveness to factors in addition to TGF-ß, which activity is subtracted out by the use of the control data obtained using the antibody treatment. Second, one can correct for spurious induction or inhibition of a TGF-ß response element by factors other than TGF-ß. The analysis of comparative data (comparing) produced by conducting the present method both with and without anti-TGF-ß antibody for the purpose of determining the level of TGF-ß in a liquid sample, can be conducted by a variety of statistical methods that are not to be construed as limiting to the invention. Exemplary comparative analyses are described in the Examples.

Contemplated for use with any of the above TGF-ß assay methods of this invention are plasmids having identifying characteristics of plasmids on deposit with ATCC having the ATCC Accession Numbers 75627, 75628 and 75629. Also contemplated are eucaryotic cells that contain the TGF-ß response element having the nucleotide sequence in SEQ ID NO 11 where the cells correspond to cells on deposit with ATCC having the ATCC Accession Number CRL 11508. In one embodiment, the use of stably transformed eucaryotic cells are contemplated.

The invention describes plasmids for use in the methods that comprise a nucleotide sequence corresponding to nucleotide sequences listed in SEQ ID NOs 1-10. TGF-ß inducible response elements that comprise a nucleotide sequence corresponding to nucleotide sequences listed in SEQ ID NOs 11-17 are also described. Contemplated promoter nucleotide sequences are listed in SEQ ID NOs 18 and 19.

A further embodiment of the methods of the invention are eucaryotic cells that are stably transformed cells containing a plasmid having a gene encoding a selectable marker for the selection of said stably transformed cells. The invention describes such plasmids having nucleotide sequences listed in SEQ ID NOs 1-6. The invention further describes a stably transformed eucaryotic cell on deposit with ATCC having ATCC Accession Number CRL 11508 containing the TGF-ß response

element having the nucleotide sequence in SEQ ID NO 11.

An additional embodiment are eucaryotic cells that are transiently transformed cells with plasmids corresponding to the nucleotide sequences listed in SEQ ID NOs 7-10.

The use of stably transformed cells is particularly preferred because it provides uniformity and reproducibility to the cell based assay without the need for additional controls for the efficiency of transformation typically associated with methods using transient transformation. Stably transformed cells do not require the use of an internal standard for transformation efficiency, and all of the cells utilized are typically uniformly transformed. Furthermore, the methods do not require the additional step of transforming the cells transiently because the stably transformed cell line is already available.

The invention describes quantifying the amount of TGF-ß in a body fluid, in culture medium, in a tissue extract, and in the like liquid samples. A further preferred embodiment is the determination of the amount of a specific isoform of TGF-ß, specifically TGF-ß1, TGF-ß2 or TGF-ß3, in a liquid sample.

In a preferred embodiment, this invention describes the use of any eucaryotic host cell that contains a TGF-ß receptor and is capable of inducing a TGF-ß response element upon activation by TGF-ß. Exemplary assays for measuring activation by TGF-ß and induction of a TGF-ß response element are described herein and can be used to identify candidate host cells suitable for use in the present diagnostic methods. A preferred host cell is a mammalian cell. Preferred mammalian cells include mink lung epithelial (MLE) cells, particularly clone C32 from MLE cells, HeLa cells, Chinese hamster ovary (CHO) cells, Hep3B cells, GM7373 cells, NIH 3T3 cells, and the like cells.

Conditions for incubating a eucaryotic cell in the present methods are the same as general cell culture methods. Typical cell culture media for culturing and incubating eucaryotic

cells include alpha-MEM, Eagle's MEM (having non-essential amino acids), RPMI 1640 and Dulbecco's modified MEM (DMEM), all which are well known in the art. The culture medium preferably contains 0.5 to 2 % (v/v) serum, preferably a fetal calf or fetal bovine serum (FCS or FBS). Cell culture conditions include the use of cells plated at a density of about 0.8 to about 3.2×10^4 cells per well of a 96-well tissue culture plate, preferably about 1.6×10^4 cells per well. Cells are typically plated at the indicated density, and allowed to grow until they reach a confluence density of from about 70% confluent to about 1 day post-confluent, but should preferably be allowed to grow after plating for a time period sufficient for the cells to express detectable levels of TGF-ß receptor, which time period is typically about 0.5-24 hours, preferably about 1-5 hours, and preferably is about 3 hours.

After plating and culturing, the eucaryotic cells are incubated under culturing conditions with culture medium that includes a predetermined volume of a liquid sample believed to contain TGF-ß. The incubation time period is a time sufficient for any TGF-ß present in the liquid sample to interact with the eucaryotic cell TGF-ß receptor and thereby induce the TGF-ß response element and express the indicator polypeptide. The time required for the expressed indicator polypeptide to accumulate to detectable levels will vary with the choice of indicator and method of detection, and can be predetermined. However, typical incubation times for contacting the cell with the liquid sample can range from 2 to 24 hours, preferably about 6 to 22 hours, more preferably 10 to 20 hours, and particularly about 14 hours. Particularly preferred culturing and incubation conditions for use in the present methods are described in the Examples.

The detection of TGF-ß in liquid samples such as body fluid or tissue extract samples is useful in following the levels of TGF-ß in patients experiencing a variety of conditions where the TGF-ß level is important to the clinician.

For example, TGF-ß levels are significant in diseases characterized by excessive fibrosis such as hepatic fibrosis and the like, in proliferative and in conditions where there is an increase in collagen expression, and the like conditions where TGF-ß is believed to participate. In addition, there are many therapeutic uses of TGF-ß, and therefore, the present assay methods are useful for measuring the therapeutic fate of administered TGF-ß in patients being treated therapeutically with TGF-ß.

F. Diagnostic Methods and Kits

The present invention also contemplates a diagnostic system in kit form for assaying the amount of TGF-ß in a liquid sample according to the present methods. The diagnostic kit contains, in an amount sufficient for at least one assay, a eucaryotic cell of this invention useful for practicing the diagnostic methods for detection of TGF-ß.

The kit can further contain a packaging material. Packaging material can include container(s) for storage of the materials of the kit, and can include a label or instructions for use.

The kit can additionally contain an aliquot of reference TGF-ß for use in generating a standard reference curve using the methods of the invention.

Thus in preferred embodiments, a diagnostic kit includes, in an amount sufficient for at least one assay, the following: (a) packaging material; (b) eucaryotic cells contained within the packaging material, where the cells are capable of expressing an indicator molecule and containing a plasmid comprising, in the direction of transcription, a regulatory region that includes at least one TGF-ß inducible response element that is operatively linked to a promoter, and a structural region downstream of said promoter, where the TGF-ß response element is capable of inducing dose-dependent indicator molecule activity and the structural region coding

for said indicator molecule; and (c) an aliquot of TGF-ß contained within said packaging material, where the TGF-ß is used for generating a reference curve as described herein representing a measured amount of the indicator molecule produced from a known concentration of TGF-ß.

As used herein, the term "packaging material" refers to a solid matrix or material such as glass, plastic, paper, foil and the like capable of holding within fixed limits eucaryotic cells and an aliquot of TGF-ß. Thus, for example, packaging material can be a plastic vial used to contain eucaryotic cells in growth medium to which liquid samples can be added for activating the TGF-ß responsive plasmid within the cells. Packaging material can also be a glass vial in which an aliquot of TGF-ß is contained for use in generating a reference curve, the latter of which is described in Section E.

As used herein, an "aliquot" of TGF-ß refers to an amount of TGF-ß sufficient to generate a reference curve of this invention. In preferred embodiments, the aliquot of TGF-ß is provided in the form of a substantially dry powder, i.e., in lyophilized form, for subsequent reconstitution or in the form of a solution, i.e., a liquid dispersion. Preferably the amount of powdered TGF-ß is in the range of 25 nanograms (ng), more preferably 125 ng to 625 ng, and most preferably 250 ng. Preferably the amount of TGF-ß in liquid solution is in the range of 1 to 50 nanomolar (nM), more preferably 5 to 25 nM and most preferably 10 nM. Preferred serial dilutions of TGF-ß used in generating the reference curve are described in Section E. The TGF-ß provided in the kit preferably includes each of the three TGF-ß isoforms as described in Section B.

The term "indicator molecule or indicator polypeptide" as used in this invention and described in Section D1 refers to a molecule encoded by a reporter gene, the expression of which in the expression vectors of this invention, results in a detectable measurable protein, polypeptide, enzyme and the like.

In preferred embodiments, the packaging material includes a label indicating that eucaryotic cells containing TGF-ß responsive expression vectors can be used for determining the amount of TGF-ß in a liquid sample that includes the steps of (a) incubating the cells with the selected liquid sample; (b) measuring the amount of the induced indicator molecule; and (c) comparing the amount of measured indicator molecule with a reference curve. Thus, the packaging material contains a label that is a tangible expression describing the methods of this invention as described in Section E. of using plasmid-transformed eucaryotic cells for quantifying the amount of TGF-ß in a test liquid sample.

The packaging materials discussed herein in relation to the kit of this invention are those customarily utilized in kits or diagnostic systems. Such materials include glass and plastic, the latter of which include polyethylene, polypropylene and polycarbonate, bottles, vials, plastic and plastic-foil laminated envelopes and the like.

The eucaryotic cells transformed with the TGF-ß responsive expression vectors of this invention are cells that express TGF-ß receptor on their cell surface as described in Section E. All normal cells and most all neoplastic cells have cell surface membrane receptors also referred to a binding proteins for TGF-ß. For review, see Tucker et al., Proc. Natl. Acad. Sci., USA, 81:6757-6761 (1984) and Frolik et al., J. Biol. Chem., 259:10995-11000 (1984). The receptors have previously been described in Section E. Preferred cells for use with the TGF-ß assay kit include mink lung epithelial cells (MLE cells), HeLa cells, Chinese Hamster Ovary cells, Hep3B cells, GM7373 cells and NIH 3T3 cells, with the C32 clone from the mink lung epithelial cells being the most preferred cell line.

In preferred embodiments, the eucaryotic cells are transformed with the expression vector plasmids described in Section D have a nucleotide sequence that corresponds to a sequence in SEQ ID NOs 1-10. Contemplated for use in the kit

are stably and transiently transformed eucaryotic cells. As described in Section D1, for preparing stably transformed eucaryotic cells, the plasmids corresponding to SEQ ID NOs 1-6 are preferred for use. A further preferred eucaryotic cell for use in the kit is the Hep3B cell line co-transfected with p1500Luc and RSVneo for preparing stably transformed cells that have been deposited with ATCC having the ATCC Accession Number CRL 11508 and identified by the designation "LUCI". For preparing transiently transformed eucaryotic cells, the plasmids corresponding to SEQ ID NOs 7-10 are preferred for use.

In preferred embodiments, eucaryotic cells for use with the kit contain a plasmid having the identifying characteristics of a plasmid on deposit with ATCC having the Accession Numbers 75627, 74628 and 75629 as described in Section C.

The kit of this invention further includes an anti-TGF-ß antibody for use in a parallel control assay for determining the amount of indicator molecule produced other than by TGF-ß induction. Preferred anti-TGF-ß antibodies are anti-TGF-ß1, anti-TGF-ß2 or anti-TGF-ß3 monoclonal antibodies commercially available from Genzyme Corp., Cambridge, MA.

Preferred diagnostic assays accomplished with the kit performed as described herein are for the quantitation of the amount of TGF-ß in a liquid sample. A liquid sample can include an isoform of TGF-ß, specifically TGF-ß1, TGF-ß2 or TGF-ß3. A liquid sample further includes any body fluid, culture medium and a tissue extract that may contain unknown quantities of TGF-ß. Thus, the liquid sample includes the body fluids, serum, plasma, whole blood, lymph fluid, synovial fluid, follicular fluid, seminal fluid, amniotic fluid, urine, spinal fluid, saliva, sputum, tears, perspiration, mucus and the like. Culture medium includes culture supernatant, also referred to as conditioned medium, collected from cells maintained in tissue culture as described in Example 3B.

Tissue extracts also encompass extracts of cells, referred to as cellular extracts. In addition, organs such as placentas can be obtained and extracted with well known procedures to prepare placental extracts. Extracts can also be obtained of any body organ or portion thereof, tissue or cells, including normal, tumorigenic, and malignant cells. This is generally accomplished by surgical means, i.e., by biopsy samples including needle aspirates, tissue scrapings, or freshly dissected tissues and the like. Extracts are the collected samples are then prepared by means including homogenization in lysis buffers, including detergents such as NP-40, Triton X-100, and the like. Common methods include using potters, blenders, ultrasound generators, and dounce homogenizers.

EXAMPLES

The following examples relating to this invention are illustrative and should not, of course, be construed as specifically limiting the invention. Moreover, such variations of the invention, now known or later developed, which would be within the purview of one skilled in the art are to be considered to fall within the scope of the present invention hereinafter claimed.

1. Preparation of Expression Vectors Containing TGF-ß Response Elements

A. Source Cloning Vector Constructs and Preparation of Expression Vectors for Stable Transformation

Eucaryotic expression vectors having a regulatory region having at least one TGF-ß response element derived from the 5' promoter end of the human plasminogen activator inhibitor type 1 (PAI-1) gene operatively linked to a PAI-1 minimal promoter and a downstream structural region containing a gene coding for an indicator polypeptide, preferably luciferase, were prepared and designated generally as PAI/L

eukaryotic expression constructs. Operatively linking refers to the covalent joining of nucleotide sequences, preferably by conventional phosphodiester bonds, into one strand of DNA, whether in single- or double-stranded form. Moreover, the joining of nucleotide sequences results in the joining of functional elements such as response elements in regulatory regions with promoters and downstream structural regions as described herein.

The expression vector constructs of this invention were then used for preparing stably transformed cells for use in the quantitative TGF-ß assays of this invention. The expression vectors were designed to contain varying lengths and arrangements of the TGF-ß response elements from the PAI-1 promoter, a neomycin-resistance conferring gene for selection and a gene encoding an indicator polypeptide, preferably luciferase. Two starting vectors were required to prepare the expression vectors having a neomycin-resistance conferring gene. One of these starting cloning plasmid vectors, designated p19Luc, was previously described by van Zonneveld et al., Proc. Natl. Acad. Sci., USA, 85:5525-5529 (1988), the disclosure of which is hereby incorporated by reference.

1) Preparation of Cloning Vector p19Luc

The promoter-less reporter gene p19Luc plasmid was originally designed by van Zonneveld et al., Proc. Natl. Acad. Sci., USA, 85:5525-5529 (1988) to monitor promoter activity with a structural region, having the firefly luciferase gene to function as a reporter gene, fused to a SV40 splice and polyadenylation site. The p19Luc plasmid also contained a multiple cloning site preceded by two SV-40-derived polyadenylation sites. The p19Luc plasmid was constructed from pSVOAL-AΔ5', a vector described by De Wet et al., Mol. Cell. Biol., 7:725-737 (1987). The pSVOAL-AΔ5' was first linearized with Hind III and one portion of the plasmid was blunt-ended by filling in the Hind III sites with E. coli DNA polymerase I

large fragment (Klenow), ligated to phosphorylated Eco RI linkers (New England Biolabs, Beverly, MA). Two of the resulting fragments, the 621 bp fragment originally containing the 5' end of the luciferase gene and the 2718 bp fragment originally located on the 5' end of this fragment, were isolated. A second portion of the Hind III-cleaved pSVOAL-AΔ5' was ligated to a 55 bp polylinker and cleaved with Eco RI. The resulting 2831 bp fragment containing the multiple cloning site and the pBR322-derived ampicillin resistance-conferring gene was isolated. These fragments were ligated to create the circular double-stranded p19Luc plasmid that contained the three fragments in their original orientation but with the multiple cloning site in the original Hind III site.

The continuous 6170 bp sense strand, also referred to as the coding strand, nucleotide sequence of an Eco RI-linearized p19LUC vector is listed in the Sequence Listing as SEQ ID NO 21. The convention adopted for listing the nucleotide sequences of the p19Luc vector as well as all the expression vectors of this invention derived from p19Luc is to list only the sense strand of each vector with the nucleotide position 1 always beginning with the middle of the Eco RI site, specifically the first T nucleotide.

The Eco RI-linearized p19Luc vector contained the following list of elements and restriction sites beginning with the 5' middle Eco RI "T" nucleotide position 1 and extending to the 3' end of the vector ending with the middle Eco RI "A" nucleotide position 6170 (nucleotide positions as listed in SEQ ID NO 21 are indicated in parentheses): a Pst I restriction site (750-755) within the pBR322-derived ampicillin resistance-conferring gene (amp); an Acc I restriction site downstream of the amp gene (2113-2118); two tandem polyadenylation sites immediately upstream of the multiple cloning site beginning with Bam HI (2771-2776) and Hind III (2778-2783), continuing with adjacent Sph I, PstI, Hinc II/Acc I/Sal I, Xba I, Bam HI, Xma I/Sma I, Kpn I, Sst I, and ending with Eco RI (2829-2834);

the luciferase gene adjacent to the Eco RI site in which are four restriction sites, Xba I (2910-2915), Eco RI (3450-3455); Sph I (3522-3527), and Xba I (4564-4569); an SV40 splice site adjacent to the 3' end of the luciferase gene followed by a third polyadenylation site; a Bam HI restriction site (5417-5422); and lastly a Pst I restriction site (5962-5967).

For use in preparing the expression vectors of this invention, the multiple cloning site in the promoterless p19Luc plasmid described above allowed for the directional ligation of both non-TGF-ß responsive promoters and TGF-ß responsive regulator regions containing TGF-ß response elements, the latter of which comprised the regulatory region of the resultant vectors. The promoters and TGF-ß response elements and the ligation thereof to form TGF-ß expression vectors are described herein and below.

Thus, the p19Luc plasmid was used as a cloning vector for construction of all the expression vectors of this invention. The advantage of using the p19Luc and the p19Luc-derived p39Luc expression cloning vectors, the latter of which is described below, is that the vectors allow for the construction of TGF-ß responsive vectors having a selected regulatory region operatively ligated to a selected promoter. Therefore, any regulatory region of any length containing one or more TGF-ß response elements can be paired with any promoter, a non-TGF-ß responsive PAI-1 or heterologous HBV promoter as used herein but not limited to that, to prepare TGF-ß responsive expression vectors that provide for the quantitation of inducing TGF-ß.

While specific expression vector constructs having the preselected regulatory regions as described herein were prepared for use in this invention, also contemplated are expression vectors having regulatory regions with TGF-ß response elements that are either longer, shorter, tandemly arranged, reversed, permutations thereof and the like operatively ligated to a selected promoter. Moreover, in addition to the construction methods detailed herein, other

methods of preparing p19Luc-derived expression vectors having TGF-ß response elements and promoters are familiar to one of ordinary skill in the art of vector construction and are described by Ausebel, et al., In Current Protocols in Molecular Biology, Wiley and Sons, New York (1993) and by Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, 1989.

2) Preparation of Expression Vector p1500Luc

One expression vector of this invention, designated p1500Luc, was constructed from p19Luc and a cosmid containing the PAI-1 promoter in which TGF-ß response elements are located. To prepare p1500Luc, a 1547 base pair (bp) Kpn I-Eco RI fragment of the PAI-1 promoter was obtained from a cosmid containing the entire PAI-1 gene (Loskutoff et al., Biochem., 26:3763-3768 (1987), the disclosure of which is hereby incorporated by reference, and was cloned into the Kpn I and Eco RI sites of pUC19, a plasmid available from American Type Culture Collection, Rockville, MD with the ATCC Accession Number 37254, to create a vector designated pUCEK19. The fragment contained the 1442 bp TGF-ß response element (SEQ ID NO 11) from the PAI-1 promoter that corresponded to nucleotide position -1481 and extended to the nucleotide position -40 continuous with a 115 bp minimal (non-TGF-ß responsive) PAI-1 promoter sense strand sequence (SEQ ID NO 18) corresponding to nucleotide position -39 ending with an E. coli DNA polymerase filled-in Eco RI site at nucleotide position at +76 as described by Bosma et al., J. Biol. Chem., 263:9129-9141 (1988). The entire 15,867 bp PAI-1 gene sequence including significant stretches of DNA that extend into its 5'- and 3'-flanking DNA regions was described by Bosma et al., J. Biol. Chem., 263:9129-9141 (1986), and is available in the GenBank™/EMBL Data Bank with accession number(s) J03764.

To create a sensitive reporter gene system with a regulatory region having the 1442 TGF-ß response element of the

PAI-1 promoter contiguous with the minimal PAI-1 promoter, the pUCEK19 plasmid prepared above was then digested with Kpn I and Eco RI and the isolated fragment was then ligated into the multiple cloning site of a similarly digested p19Luc. The resulting vector was designated p1500Luc.

3) Preparation of Expression Vector p800Luc

Another vector, designated p800Luc, was prepared for subsequent constructon of p800neoLuc as described below. The p800Luc plasmid, having a deletion in the 5' end of the PAI-1 construct so that the 5' end began with the -800 nucleotide in the native PAI-1 promoter, was prepared by digesting the PAI-1-gene-containing cosmid described above with Hind III and Eco RI. The actual Hind III-Eco RI digest of the PAI-1 promoter resulted in a fragment that corresponded to nucleotides -799 to +71 bp in the PAI-1 promoter that was subsequently ligated into a similarly digested p19Luc vector forming a PAI-1 region extending from nucleotide -800 to +76. The resulting p800Luc plasmid retained all the features of p19Luc with the exception of the insertion of the PAI-1-derived regulatory region having a TGF-ß response element and a promoter.

The restriction fragments described to prepare p1500Luc and p800Luc had an identical 3' end (an Eco RI site at +71 nucleotide of the PAI-1 promoter) and a different 5' end. The vectors, p1500Luc and p800Luc, were used for transient transformations as they lacked a selectable marker gene. The p1500Luc plasmid was also used to prepare stable transformations with a second vector as described in Example 1C. In addition, the p800Luc served as the starting cloning construct for the preparation of p800neoLuc as described below. The TGF-ß response element in the -800 to +76 PAI-1 promoter region began at -800 and ended at -40, the nucleotide sequence of which is listed in SEQ ID NO 12. The remaining nucleotides comprised the non-TGF-ß responsive minimal promoter in this

PAI-1 fragment are listed in SEQ ID NO 18.

4) Preparation of Cloning Vector p39Luc

An expression vector, designated p39Luc, having a promoter for activating transcription of the luciferase gene while lacking TGF-ß response elements, thereby lacking responsiveness to TGF-ß, was prepared as described by Keeton et al., J. Biol. Chem., 266:23048-23052 (1991). A fragment of the PAI-1 promoter (i.e., between -39 and +76, which had been determined in the TGF-ß assay as described in Example 3A to have low basal activity and only minimal response to TGF-ß (average induction of 2.7-fold), was used as a minimal promoter in the constructs for use in quantifying the amount of TGF-ß in a test liquid sample. Since the minimal promoter sequence conferred only a minimal background response to TGF-ß as shown in Example 3A, the minimal PAI-1-derived promoter is also referred to as being "non-TGF-ß responsive".

Briefly, the p800Luc vector was linearized by digestion with Hind III followed by 5' digestion of PAI-1 promoter with Bal-31 slow exonuclease (International Biotechnologies, New Haven, CT) as described by Keeton et al., J. Biol. Chem., 266:23048-23052 (1991). The digestion was allowed to proceed until the -39 nucleotide position of the PAI-1 promoter was reached. Thereafter, the linearized and Bal-31 digested plasmid was ligated with T4 ligase forming a double-stranded circular vector designated p39Luc.

The resultant expression vector, into which TGF-ß response elements were subsequently ligated as described in Example 1C, contained the PAI-1 minimal promoter nucleotide sequence corresponding to -39 to +76 of the promoter as listed in SEQ ID NO 18. This minimal promoter was operatively linked to and continuous with the structural region that contained the firefly luciferase gene present in the vector. Since the p39Luc cloning vector was derived from p800Luc which itself was derived from p19Luc, the remaining elements and features of the

vector were retained unchanged from p19Luc. The 6229 bp sense strand nucleotide sequence of the Eco RI-linearized p39Luc vector is listed in the SEQ ID NO 23.

The p39Luc cloning expression vector is also obtained by preparing a double-stranded oligonucleotide sequence corresponding to the sequence in SEQ ID NO 18 and ligating it into the Hind III/Eco RI multiple cloning site of p19Luc. The overhang from the Hind III/Eco RI digests in the p19Luc vector is first digested with mung bean nuclease and followed by ligation with the blunt-ended double-stranded oligonucleotide promoter. Other construction methods are well known to and easily accomplished by one of ordinary skill in the art.

The p39Luc vector was useful for operatively ligating regulatory regions that contained TGF-ß response elements resulting in an expression vector that was responsive to DNA-binding proteins, the result of which was induction of the transcription and translation of the indicator molecule, luciferase. TGF-ß responsive expression vectors for use in practicing this invention having TGF-ß response elements other than those specified herein are readily constructed through the use of either p19Luc or p39Luc starting cloning expression vectors.

5) Preparation of Cloning Vector HBVLuc

To create expression vectors having heterologous non-TGF-ß responsive promoters instead of having the PAI-1-derived minimal promoter described above, a minimal promoter construct derived from the Hepatitis B viral promoter (HBV) was selected. This promoter contained the nucleotide sequence from -188 to +145 of the Hepatitis B promoter and showed only a 4-fold induction in response to TGF-ß. The sense strand of the double-stranded nucleotide sequence of the HBV minimal promoter is listed in SEQ ID NO 19. This promoter corresponded to the nucleotide sequence from -188 to +145 of the Hepatitis B promoter and showed only 4-fold induction in response to TGF-ß.

The 6464 bp sense strand nucleotide sequence of the Eco RI-linearized pHBVLuc vector is listed in the SEQ ID NO 25.

6) Preparation of Expression Vector p800neoLuc

For preparing an expression vector for use in stable transformations, the neomycin-resistance conferring gene from pMAMneo (Clontech, Palo Alto, CA) was inserted into the p800Luc vector containing -800 to +76 of the 5' end of the human PAI-1 gene followed by the firefly luciferase gene. As shown in Figure 1, p800Luc prepared above was first digested with Acc I, repaired to blunt ends with the Klenow fragment of DNA polymerase I, and then was isolated. The pMAMneo plasmid was digested with Sal I and Eco RI then blunt-ended with Klenow. The neomycin-resistance gene containing fragment was then isolated and had the 4302 bp sense strand nucleotide sequence listed in the Sequence Listing in SEQ ID NO 20. The linearized p800Luc and neomycin-resistance fragment were ligated, and one clone with the insert in the correct orientation was selected by restriction mapping and designated p800neoLuc. The entire Eco RI-linearized 11293 bp nucleotide sequence of the sense strand of the double-stranded p800neoLuc vector is listed in the Sequence Listing in SEQ ID NO 1. DNA sequencing was performed by a modification of the dideoxy chain-termination procedure with a Sequenase kit (United States Biochemical; Cleveland, OH). This clone, purified from large scale plasmid preparations via $CsCl_2$ gradients, was used for subsequent transfections.

Since the p800neoLuc cloning vector was derived from p800Luc which itself was derived from p19Luc, the remaining elements and features of the vector were retained unchanged from p19Luc. The p800neoLuc vector thus contained the neomycin-resistance conferring gene providing for stable transformants. The p800neoLuc vector also contained an operatively ligated regulatory region that contained TGF-ß

response element in the sequence corresponding to -800 to -40 of the PAI-1 promoter resulting in an expression vector that was responsive to TGF-ß. With this expression vector construct, the induced activation of the transcription and translation of the indicator molecule, luciferase, was obtained further allowing for the quantitation of the amount of TGF-ß responsible for activating gene expression.

7) Preparation of Cloning Vector p39neoLuc

To create an expression vector useful for constructing TGF-ß responsive vectors that resulted in stably transformed cells, the p39Luc cloning vector prepared above was linearized as described above for p800Luc and ligated with the neomycin-resistance conferring gene fragment from pMAMneo. The construction of the vector was performed as described in Example 1A6). The resultant p39neoLuc cloning expression vector had the Eco RI-linearized 10533 bp sense strand nucleotide sequence listed in the SEQ ID NO 22. Regulatory regions containing TGF-ß response elements were operatively ligated 5' to the minimal promoter sequence of the p39neoLuc as described in Example 1C for the preparation of plasmids for transient transformation.

8) Preparation of Cloning Vector pHBVneoLuc

To create an expression vector useful for constructing TGF-ß responsive vectors with a heterologous promoter for stably transforming cells, the pHBVLuc cloning vector prepared above was linearized as described above for p800Luc and ligated with the neomycin-resistance conferring gene fragment from pMAMneo. The construction of the vector was performed as described in Example 1A6). The resultant pHBVneoLuc cloning expression vector had the Eco RI-linearized 10768 bp sense strand nucleotide sequence listed in the SEQ ID NO 24. Regulatory regions containing TGF-ß response elements were operatively ligated 5' to the minimal promoter sequence of

the pHBVneoLuc as described in Example 1C for preparing plasmids for transient transformation.

9) Preparation of p1500neoLuc, p800/636neoLuc, p56neoLuc, p674neoLuc, p743neoLuc and p732neoLuc Expression Vectors

The p1500Luc vector prepared above is similarly ligated with the neomycin-resistance gene from pMAMneo to form p1500neoLuc. Other PAI-1-promoter containing expression vectors lacking the neomycin resistance gene, p800/636Luc, p56Luc, p674Luc, p743Luc and p732Luc, containing smaller TGF-ß response elements were prepared as described in Example 1C. To create the corresponding neomycin-resistance expression vectors for stably transforming recipient cells, the neomycin-resistance gene from pMAMneo is separately ligated with each of these five vectors to form expression vectors used for generating stable cell transformations. The five resultant vectors having the neomycin-resistance gene inserted are designated p800/636neoLuc (10697 bp), p56neoLuc (10549 bp), p674neoLuc (10558 bp), p743neoLuc (10569 bp) and p732neoLuc (10558 bp) and have the respective complete nucleotide sequences of the sense strand from the Eco RI-linearized double-stranded vectors in SEQ ID NOs 2-6.

Depending on the vector into which the PAI-1 promoter fragments were cloned, the designated names either had "Luc" alone or "neoLuc" respectively for vectors lacking the neomycin (neo) selectable marker gene or containing it. In addition, the plasmids were further designated by the 5' end of the PAI-1 TGF-ß response element. For example, five plasmids with shorter TGF-ß response elements were thus named p800/636neoLuc, p56Luc, p674Luc, p743Luc and p732Luc.

As with all the expression vectors of this invention, the operative elements from the original cloning vector p19Luc, from which the vectors were all derived, were retained.

The above neomycin-resistance containing expression vectors were then used in the TGF-ß assay method as described in Example 3 following transformation of host recipient cells.

B. Expression Vectors for Co-Transformation of TGF-ß Responsive Vectors and a Selectable Marker Vector for Stable Transformation

Stably transformed Hep3B cells were also obtained as described in Example 2B below through the use of co-transfections of a TGF-ß responsive vector lacking a selectable marker gene of this invention, specifically the p1500Luc prepared in Example 1A3), with a selectable marker vector, RSVneo, available from American Type Culture Collection (ATCC), Rockville, MD, ATCC Accession Number 37198. The stably transformed cell line containing plasmid p1500Luc, designated LUCI, was deposited with the ATCC on or before December 16, 1993 and was assigned the ATCC Accession Number CRL 11508.

C. Expression Vectors for Transient Transformation

Additional TGF-ß responsive expression vectors were prepared for use in this invention. In the vectors prepared as described herein, the TGF-ß response elements having a smaller length, thereby providing responsiveness to TGF-ß with reduced or absent responsiveness to other growth modulators, were made by either restriction digestion of the PAI-1 promoter or synthesizing double-stranded blunt-end oligonucleotides. The oligonucleotide sequences corresponded to preselected regions of the PAI-1 promoter sequence. The resultant TGF-ß response elements present within a regulatory region were then directionally ligated into p39Luc or p39HBV.

The regulatory region from the PAI-1 promoter corresponding to nucleotide position -800 up to and including -636 was obtained by restriction digestion and had the following sense strand sequence:

5'AAGCTTACCATGGTAACCCCTGGTCCCGTTCAGCCACCACCACCCCACCCAGCACACCTCC

AACCTCAGCCAGACAAGGTTGTTGACACAAGAGAGCCCTCAGGGGCACAGAGAGAGTCTGGAC ACGTGGGGAGTCAGCCGTGTATCATCGGAGGCGGCCGGGCA3' (SEQ ID NO 13). The additional selected regions for preparing oligonucleotides included the following sense strand nucleotide sequences with the indicated nucleotide positions as present in the intact PAI-1 promoter: 1) promoter nucleotide position -56 up to and including -41: 5'AGTTCATCTATTTCCT3' (SEQ ID NO 14); 3) promoter nucleotide position -674 up to and including -650: 5'GTGGGGAGTCAGCCGTGTATCATCG3' (SEQ ID NO 15); 4) nucleotide position -743 up to and including -708: 5'CTCCAACCTCAGCCAGACAAGGTTGTTGACACAAGA3' (SEQ ID NO 16); and 5) nucleotide position -732 up to and including -708: 5'GCCAGACAAGGTTGTTGACACAAGA3' (SEQ ID NO 17). The complementary sequences to each of the sense oligonucleotide sequences were also synthesized to allow for the formation of double-stranded oligonucleotides for ligation 5' to the PAI-1 minimal promoter sequence containing the TATA box.

The resulting double-stranded oligonucleotides were then separately operatively linked to the -39 position of this minimal promoter sense strand sequence listed in SEQ ID NO 18 present in the expression vector, p39Luc, prepared as described in Example 1A4). The sequences were confirmed by double-stranded sequencing methods.

The resulting five plasmids with shorter TGF-ß response elements were thus named p800/636Luc, p56Luc, p674Luc, p743Luc and p732Luc. The plasmids, p56Luc, p674Luc, p743Luc and p732Luc, have the respective complete sense strand nucleotide sequences beginning with the middle T of the Eco RI site as previously described listed in SEQ ID NOs 7-10. The plasmids, p674Luc, p743Luc and p732Luc, were deposited with ATCC as described in Example 5 and respectively assigned the ATCC Accession Numbers 75627, 75628 and 75629.

In similar procedures, five plasmids having a heterologous hepatitis B viral promoter, HBV, instead of the PAI-1 minimal promoter were prepared with the shorter TGF-ß response

elements, p800/636Luc, p56Luc, p674Luc, p743Luc and p732Luc. The HBVLuc cloning expression vector was prepared as described in Example 1A4). The TGF-ß response elements were ligated into linearized HBVLuc, prepared as described in Example 1A5), to form TGF-ß response element-containing plasmids lacking the neomycin-resistance-conferring gene.

Furthermore, as previously mentioned, the cloning vector constructs, p19Luc and p39Luc, provide for the operative linking of preselected regulatory regions with preselected promoters, both of which are not limited to the specific constructs described herein and above. Additional TGF-ß response elements in varied lengths and arrangements along with promoters that provide for the transcription of the reporter gene are contemplated for use in this invention.

2. Transformation of Eucaryotic Cells with Expression Vectors Containing TGF-ß Response Elements

A. Recipient Eucaryotic Cells

To identify the cell types most responsive to TGF-ß in which to transfect the TGF-ß responsive expression vectors for use in assaying the amount of TGF-ß, the vectors prepared in Example 1 were transfected as described in Example 2B and 2C into recipient cell lines including mink lung epithelial cells (MLE cells) (ATCC CCL 64), HeLa cells (ATCC CCL 2), Chinese hamster ovary (CHO cells) (ATCC CCL 61), GM7373 (chemically transformed metal bovine aortic endothelial cells or BAEs) (NIGMS Human Genetic Mutant Cell Repository, Camden, NJ), Hep3B (ATCC HB 8064) and NIH 3T3 cells (ATCC CRL 1658).

B. Stable Transformation

For preparing stably transfected cells for use with expression vectors containing the pMAMneo construct prepared in Example 1A, transfections of mink lung epithelial cells (hereinafter referred to as MLE cells to distinguish from the TGF-ß proliferation assay called MLEC) were performed. The MLE

cells were seeded at 7 x 10^5 cells/100 mm dish for 24 hours at which point they were transfected with the PAI/L construct, p800neoLuc, by calcium phosphate precipitation as described by Wigler et al., Proc. Natl. Acad. Sci., USA, 76:1373-1376 (1979). Twenty-four hours after transfection, the medium was replaced and supplemented with 400 µg/ml of Geneticin. The resistant cells were expanded in mass culture or cloned by limiting dilution for further experiments. Following selection, transfected MLE cells were maintained in DMEM containing 10% fetal calf serum and 250 µg/ml Geneticin (G-418 sulfate) (Gibco BRL, Grand Island, NY).

Stable transformations are also performed as described above with the expression vectors, p800/636neoLuc, p56neoLuc, p674neoLuc, p743neoLuc and with p732neoLuc, all of which are prepared as described in Example 1A.

C. Stable Transformation Obtained by Co-transfection of Cells

For transfecting 6 wells, 15 micrograms (µg) of p1500Luc expression vector prepared in Example 1A2) that did not have a neomycin-resistance gene was admixed with 3 µg of a plasmid encoding the neomycin selectable marker gene driven from a respiratory syncytial virus promoter, RSVneo. The RSVneo plasmid is available from ATCC with ATCC Accession Number 37198. Hep3B cells at a concentration of 6 X 10^5 cells/well were seeded as described above in Example 1B for 24 hours at which point they were transfected with the PAI/L construct, p1500Luc, by calcium phosphate precipitation followed by selection with Geneticin. The resultant cell line stably transformed with p1500Luc, designated LUCI, was deposited with ATCC on December 16, 1993 and was assigned the ATCC Accession Number CRL 11508.

D. Transient Transformation

For preparing transiently transformed cells

containing TGF-ß responsive expression vectors lacking the neomycin resistance gene prepared as described in Example 1C, Hep3B human hepatoma cells obtained from ATCC (ATCC Accession Number HB8064) were maintained in DMEM/HAMs F-12 (Whittaker Bioproducts, Walkersville, MD) supplemented with 10% fetal bovine serum (Hyclone Laboratories, Logan, UT), glutamine, sodium pyruvate, non-essential amino acids and penicillin/streptomycin (Whittaker). For transfection experiments, semiconfluent cells in 6-well (10 cm^2 per well) tissue culture plates (Corning Inc., Corning, NY) were washed twice with serum free media (DMEM/F-12) then incubated in serum free media. Separate mixtures (50 ul/well) of lipofectin (GIBCO, Grand Island, NY) at a concentration of 12.5 µg/well and DNA vector constructs prepared in Example 1A-1C at a concentration of 2.5 µg/well each in water were added to the cell-containing wells and the plates were incubated for 18 hours. After lipofection, plates were incubated an additional 24 hours in the absence or presence of 1 ng/ml TGF-ß provided by Berlix Biosciences, South San Francisco, CA. The monolayers were then washed followed by extraction into 0.25% Triton X-100. Each construct was tested with at least 2 independent DNA preparations in order to rule out any effects related to differences in DNA preparation. For each experiment, two independent transfections were performed with every construct.

3. Method for Quantifying the Amount of TGF-ß in a Sample

A. The TGF-ß Assay Method

The p800neoLuc construct stably transfected into Hep3B cells was used in the initial characterization of the assay method as described herein. TGF-ß measurement assays performed with cells transiently transformed with the remaining expression vectors containing TGF-ß response elements are presented in Example 4.

The TGF-ß assay allows for the quantification of the

amount of TGF-ß in a liquid sample, either containing purified TGF-ß or TGF-ß in a heterogeous admixture. The assay system provides for the quantification of TGF-ß through the expression of an indicator polypeptide, such as luciferase. When TGF-ß receptor-bearing cells, transfected with a TGF-ß responsive expression vector of this invention, are exposed to TGF-ß, the activation of the TGF-ß response element in the vector results in the concomitant expression of luciferase. The resulting expressed luciferase is isolated then measured as described herein. The measured luciferase resulting from activation by TGF-ß in the test liquid sample is then compared to a standardized reference curve.

This reference curve is obtained from parallel assays performed by exposing similarly transfected cells to a range of known measured amounts of TGF-ß, one or more of the known TGF-ß isoforms. The resulting expressed luciferase is then determined in a luminometer. A reference curve is then generated by plotting the measured amount of expressed luciferase against the known range of inducing amounts of TGF-ß. The amount of unknown TGF-ß in the test liquid sample is then determined by extrapolating the measured amount of test luciferase to the reference curve. The use of standard curves in quantifying the amount of protein in a liquid sample in general has been described by Lowry et al., J. Biol. Chem., 193:265-275 (1951), the disclosure of which is hereby incorporated by reference. As shown in the Examples herein, the TGF-ß assay of this invention allows for the measurement of TGF-ß from the expression and subsequent detection of an indicator polypeptide from a concentration range from less than 5 picograms/ml (pg/ml) equivalent to 0.2 pM to 10 ng/ml equivalent to 0.4 nM. The dose-dependent response is linear between 0.2 pM up to 30 pM and even up to 100 pM depending on the assay conditions.

An additional aspect of the assay for quantifying TGF-ß in complex solutions was the use of neutralizing anti-TGF-ß

monoclonal antibodies admixed with the test liquid sample in assays run in parallel to untreated test liquid samples as described in Example 3B. These control assays are used to determine if other molecules are present in the test sample that can affect the assay through either inhibition or activation of other regions of the truncated PAI-1 promoter. For example, conditioned medium obtained from cell cultures and body fluids contain growth factors and DNA binding proteins that function as transcriptional activators or inhibitors. If a corresponding response element for an additional non-TGF-ß activator or inhibitor is present in the expression vector, the binding of that molecule to the response element may cause enhanced or diminished expression of the indicator polypeptide. By antibody neutralization of the TGF-ß in the test sample, any residual measured luciferase can then be ascribed to non-TGF-ß activation.

The shorter TGF-ß response elements used in the expression vector systems of this invention, even including the longer p800neoLuc, are less likely to have non-TGF-ß response elements that are bound by other DNA-binding proteins as shown in Examples 3C-3F. Thus, the use of parallel antibody control assays to allow for a determination of the amount of luciferase produced from only TGF-ß activation is preferred when expression vectors having longer response elements are used. Moreover, while the TGF-ß assay is not isoform specific, using the appropriate standard reference curves and parallel assays with neutralizing antibodies to the various TGF-ß species allows for quantification of unique TGF-ß isoforms.

In the assays described herein, the various following reagents including their sources are listed: recombinant human TGF-ß1 (rTGF-ß1) (gift from Berlix Biosciences, South San Francisco, CA); rTGF-ß2 and neutralizing monoclonal antibodies against TGF-ß1, TGF-ß2 and TGF-ß3 (Genzyme, Cambridge, MA); rTGF-ß3, recombinant human interleukin-1alpha (rIL-1alpha) and recombinant human platelet-derived growth factor-BB (PDGF-BB)

(R&D Systems, Minneapolis, MN); recombinant human basic fibroblast growth factor (bFGF) (Synergen Inc., Boulder, CO); epidermal growth factor (EGF) from mouse submaxillary glands (Boehringer Mannheim Biochemicals, Indianapolis, IN); dexamethasone, retinoic acid, and plasmin (Sigma Chemical Co., St. Louis, MO); thrombin (Armour Pharmaceutical Co., Kankakee, IL); and hematopoetic factors granulocyte-colony stimulating factor (GCSF), granulocyte-macrophage-colony stimulating factor (GMCSF), stem cell factor, and IL-3 (Amgen, Thousand Oaks, CA).

The TGF-ß quantification assay of this invention was performed as follows: 1.6×10^4 stably transfected MLE cells per well plated in 96 well tissue culture dishes were allowed to attach for 3 hours at 37°C in a 5% CO_2 incubator. The medium was replaced with the test sample containing unknown quantities of TGF-ß, DMEM, 0.1% BSA (DMEM-BSA) containing rTGF-ß1, rTGF-ß2, rTGF-ß3, IL-1alpha, PDGF-BB, bFGF, or EGF for 14 hours at 37°C. Time courses of exposure to the samples were performed as shown for optimizing the assay as shown below. However, in general, approximately 24 hours after additions of the sample to the transfected cells, the cells were observed under phase contrast microscopy. At least in one vector-transfected cell line, Hep3B cells, the presence of TGF-ß in quantities at least or greater than 0.1 ng/ml TGF-ß in the sample was detected visually by the change of morphology and density of the cell population. The untreated cells remained organized with cell size decreasing upon confluence until the cell borders were no longer visible. In the presence of TGF-ß, the untreated cell density was never attained and the cells were larger, flatter and less organized.

Following visual inspection, cell extracts were prepared and assayed for luciferase activity using the enhanced luciferase assay kit (Analytical Luminescence, San Diego, CA) as per the manufacturer's illustructions. Treated cells were first washed twice with 2 ml phosphate-buffered saline (PBS) without Ca^{++} and Mg^{++} and then extracted with 100 ul of 0.25%

Triton-X 100 (cell lysis buffer, Analytical Luminescence). The plates were gently shaken until the monolayer detached from the plastic. The plates were then placed on a rotator at room temperature for 20 minutes.

Eighty ul of the resultant lysates were transferred to a Microlight 1 96-well plate (Dynatech Laboratories Inc., Chantilly, VA) and were analyzed using an ML1000 luminometer (Dynatech) with 100 ul injections of both Substrates A and B (Analytical Luminescence). Luciferase activity was reported as relative light units (RLU) as measured by the light generated over a ten second period. All assays were performed in triplicate. Error bars in the collected data represent the standard error of the mean of the samples.

To quantitate the amount of TGF-ß inducing the measured amount of luciferase from liquid samples, reference curves were prepared from parallel assays performed by exposing similarly transfected cells to a range of known measured amounts of TGF-ß, one or more of the known TGF-ß isoforms. Serial dilutions of the control TGF-ß concentrations were prepared from a 1 nanomolar (nM) concentration down to 0.078 picomolar (pM). The TGF-ß assay was performed for each serial dilution and the resulting expressed luciferase was then determined in a luminometer. A reference (standard) curve was then generated by plotting the measured amount of expressed luciferase against each of the known concentrations of inducing amounts of TGF-ß. The amount of unknown TGF-ß in the test liquid sample was then determined by extrapolating the measured amount of test luciferase to the reference curve.

B. Sensitivity of the TGF-ß Assay Method

To identify the cell type most responsive to TGF-ß for use in the methods of this invention, the p800neoLuc construct prepared in Example 1A was stably transfected as described in Example 2B into a variety of cell lines including MLE cells, HeLa, Chinese hamster ovary (CHO), GM7373 cells

(chemically transformed fetal bovine aortic endothelial cells obtained from the NIGMS Human Genetic Mutant Cell Repository, Camden, NJ) and NIH 3T3 cells. After treatment of the transfected cell lines with recombinantly-produced TGF-ß1, designated rTGF-ß1, the cell lysates were assayed for luciferase activity and protein content. There was a linear relationship between the luciferase activity and the protein content of the cell lysates between 0.7 and 14 µg for all of the cell lines. Nontransfected parental cells demonstrated no detectable luciferase activity. Of the various cell lines, the transfected MLE cells demonstrated the greatest sensitivity to TGF-ß. After cloning the transfected MLE cells by limiting dilution, cells from clone 32 (C32) were found to be the most sensitive and were used for all subsequent assays.

C32 cells were sensitive to rTGF-ß1, ß2 and ß3 in the picomolar (pM) to the nanomolar (nM) range as evidenced by increased luciferase activity in relative light units (RLU) as shown in Figure 2A. All three isoforms, rTGF-ß1, rTGF-ß2 and rTGF-ß3, respectively graphed as closed squares, closed circles and closed triangles, demonstrated good dose dependant responses particularly at low TGF-ß concentrations (<4 pM: 100 pg/ml) where the responses were essentially linear (Figure 2B). rTGF-ß3 was the most potent inducer of luciferase activity consistent with the observation that MLE cells were most sensitive to this isoform of TGF-ß3 as described by van Zonneveld et al., Proc. Natl. Acad. Sci., USA, 85:5525-5529 (1988) (see also Figure 6 as described in Example 3E).

To further assess the dose-dependent responsiveness of luciferase activity by TGF-ß induction, the TGF-ß assay was performed with 8 pM of rTGF-ß1, rTGF-ß2 or rTGF-ß3 in DMEM-BSA in the presence (partially filled squares) or absence (open squares) of 100 µg/ml of anti-TGF-ß1, anti-TGF-ß2 or anti-TGF-ß3 monoclonal antibodies (Genzyme Corp., Cambridge, MA). As shown in Figure 2C, the induction of luciferase activity by rTGF-ß1, rTGF-ß2 and rTGF-ß3 was inhibited by the addition of

rTGF-ß1, rTGF-ß2 and rTGF-ß3 neutralizing monoclonal antibodies as compared to the baseline induction obtained when using medium alone (filled squares).

The effects of cell culture medium, cell density and assay incubation time on the sensitivity of the TGF-ß assay was also assessed. To test the effects of cell culture medium, the TGF-ß assay was performed using increasing concentration of rTGF-ß1 in DMEM (closed squares), alpha-MEM (closed circles), CMEM (Eagles medium supplemented with nonessential amino acids; closed triangles), or RPMI-1640 (closed diamonds). All media contained 0.1% BSA. The quantification of TGF-ß in test samples was accomplised in the TGF-ß assay in all tested media as shown in Figure 3A, although samples assayed in DMEM yielded the greatest luciferase activity.

The effect of different cell plating densities on the induction of luciferase activity by rTGF-ß1 were also examined when transfected cells were maintained in the presence of DMEM. For this assay, increasing concentrations of rTGF-ß1 in DMEM and 0.1% BSA were measured using 3.2×10^4 (closed squares), 1.6×10^4 (closed circles), or 0.2×10^4 (closed triangles) C32 cells/well after a three hour attachment period. The test samples were maintained with the transfected cells for 14 hours prior to assaying for luciferase activity. The results graphed in Figure 3B show that 1.6×10^4 cells/well were found to yield the best overall results. Cell densities greater than 1.6×10^4 cells/well decreased the sensitivity of the assay at low TGF-ß concentrations and did not significantly increase sensitivity at higher TGF-ß3 levels. Decreasing the concentration of cells to 0.8×10^4 cells/well increased the sensitivity at low TGF-ß3 levels (Figure 3D (inset in Figure 3C) but decreased sensitivity at higher TGF-ß concentrations.

Unlike the traditional MLEC assay where the density of the cells prior to plating affects the sensitivity, there was little or no difference whether the cells were 70% confluent, confluent or 1 day post confluent prior to plating for the TGF-

ß assay. The cell attachment and incubation times, however, did affect the sensitivity. When C32 cells were plated for 2, 3 or 4 hours prior to the addition of samples, a 3 hour plating time appeared to be optimal. Shorter plating times decreased sensitivity, whereas longer times had little effect on the subsequent assay.

Incubation time with the sample also affected the assay. After a three hour attachment period, 1.6 X 10^4 C32 cells were incubated with various concentrations of rTGF-ß1 ranging from 0 to 50 pM for 6 (closed squares), 14 (closed circles) or 22 hours (closed triangles) prior to assaying for luciferase activity as shown in Figure 3C. Incubation times of 12-14 hours were found to give the best results over the widest concentration range. The sensitivity of cells incubated for 6 hours was not as great at higher TGF-ß1 concentrations, whereas the sensitivity of cells incubated for 22 hours was decreased at low TGF-ß1 concentrations. There also appeared to be a slight decrease in sensitivity to TGF-ß as the cells were repeatedly passaged (>30). This phenomenon was observed for the MLEC assay as well.

C. Specificity of the TGF-ß Assay Method

After examining the sensitivity of the assay, specificity of the TGF-ß assay was then examined. Four known inducers of PAI-1 expression, were incubated with C32 cells and the luciferase activity determined. The inducers tested included fibroblast growth factor (bFGF) (Saksela et al, J. Cell Biol., 105:957-963 (1987)), platelet-derived growth factor (PDGF-BB) (Reilly et al., J. Biol. Chem., 266:9419-9427 (1991)), interleukin-1 alpha (rIL-1alpha) (Schleef et al., J. Biol. Chem., 263:5797-5803 (1988)) and epidermal growth factor (EGF) (Seebacher et al., Exp. Cell Res., 203:504-507 (1992) and Sato et al., Exp. Cell Res., 204:223-229 (1993)). The assay was performed as described in Example 3A with DMEM-BSA containing rTGF-ß1 (closed squares), recombinant human bFGF

(closed circles), recombinant IL-1alpha (closed triangles), recombinant PDGF-BB (closed triangles) or EGF (open squares) ranging in concentration from 0.1 to 500 pM. As seen in Figure 4A, even at high concentrations of these factors (500 pM), there was little or no induction of luciferase expression except by PDGF which demonstrated a slight induction.

Additional inducers of PAI-1, dexamethasone (10^{-7} M), retinoic acid (1 uM), plasmin (0.1 U/ml), thrombin (1 U/ml), and the hematopoetic factors granulocyte colony stimulating factor (10 ng/ml; 525 pM), granulocyte-macrophage-colony stimulating factor (10 ng/ml; 690 pM), stem cell factor (50 ng/ml; 2.7 nM) and IL-3 (10 ng/ml; 666 pM), were also tested for their ability to induce luciferase expression in the assay method of this invention. Only plasmin and thrombin elicited minor elevations of luciferase activity that were inhibited by the addition of aprotinin or hirudin, respectively. Of the molecules tested in the TGF-ß cell assay, only the TGF-ßs demonstrated dose-dependent increases in luciferase expression.

When these factors were tested in the presence of TGF-ß1, a slightly different pattern emerged. These assays were performed with C32 cells maintained in DMEM/BSA containing 1 pM rTGF-ß1 (closed squares) separately admixed with each of the growth factors, bFGF (closed circles), recombinant IL-1alpha (closed triangles), recombinant PDGF (closed diamonds) or EGF (open squares), ranging in concentration from 0.2 to 500 pM. The results, graphed in Figure 4B, show that high concentrations (500 pM) of PDGF-BB and rIL-1alpha increased the luciferase ativity above that induced by TGF-ß alone. bFGF had a similar effect that was observed at lower concentrations. This induction, maximal at 10 pM bFGF, was abrogated by the addition of bFGF neutralizing antibodies, and did not increase at higher concentrations (>10 nM) of bFGF.

Because this enhancement may have resulted from a bFGF-mediated increase in total cell number and/or protein, crystal violet staining of parallel cultures and protein assays of the

cell lysates was performed. The normalization of the amount of protein using these values, however, did not reduce the luciferase activity in the bFGF plus rTGF-ß1-treated cultures to that of cells treated with rTGF-ß1 alone. Interestingly, uncloned transfected MLE cells were less sensitive to bFGF and other factors including TGF-ß.

Additional TGF-ß assays were performed using the ATCC deposited LUCI cell line containing the p1500Luc expression vector co-transfected with RSVneo as described in Example 2C to determine the specificity of activation of the PAI-1 promoter by other cell activating molecules (agents). The TGF-ß assays were performed as described in Example 3A with the exception that the p1500Luc vector was used instead of the p800neoLuc vector. Controls in these assays included the use of two additional luciferase-expressing vectors that had the vitronectin (VN) and respiratory synctial virus (RSV) promoters in place of the PAI-1 truncated promoter. The molecules used in the assays included the following: (the source and concentrations are indicated in the parentheses) 1) human recombinant IL-6 (Boerhringer Mannheim, Indianapolis, IN; 500 U/ml); 2) dexamethasone (Sigma Chemical Co.; 10^{-5}M); 3) TGFß-ß (Berlix Biosciences; 1 ng/ml); 4) lipopolysaccharide (LPS) (Sigma Chemical Co.; 1 ng/ml); 5) human recombinant alpha tumor necrosis factor (TNF) (Boehringer Mannheim; 100 ng/ml); 6) human recombinant IL-1 (Sigma Chemical Co.; 50 U/ml); and 7) thrombin (NY State Department of Health, Albany, NY; 10 U/ml).

The assays were performed as indicated in Table 1 in which the fold induction is indicated as measured by relative light units of luciferase that resulted from the activation of either the PAI-1, VN or RSV promoters when exposed to the various agents.

Table 1

Agents	PAI-1	VN	RSV
Control	1X	1X	1X
IL-6	2X	15X	1X
Dexamethasone	1X	1X	1X
Il-6 + Dex.	6X	26X	2X
TGF-ß	147X	1X	2X
LPS	2X	1X	1X
TNF	0.7X	0.3X	0.8X
IL-1	0.9X	0.3X	1X
Thrombin	1X	0.9X	1X

The 1500 bp PAI-1 promoter present in the p1500Luc vector was slightly responsive to IL-6, LPS and a mixture of IL-6 plus dexamethasome. In contrast, the induction of luciferase expressing in response to activation by TGF-ß was 147-fold over that seen in the control untreated cells. Furthermore, IL-6 and IL-6 plus dexamethasone were effective activating agents when used in the presence of a vitronectin promoter. None of the agents were significantly effective at inducing expression from the RSV promoter.

These results confirm that TGF-ß is the predominant activator of the PAI-1 promoter and that the TGF-ß assay of this invention exhibits remarkable specificity. Thus, the assay is valuable in that the measurement of TGF-ß that has been purified or even TGF-ß present in unknown quantities in a complex solution containing many promoter-specific molecules can be readily determined without confounding by contaminants. With the added control of pre-treating the liquid samples with neutralizing antibodies to TGF-ß isomers, the absolute amounts of TGF-ß as well as isomer type can be determined.

D. Effects of Serum for Quantifying TGF-ß in the TGF-ß Assay Method

To assess the effects of serum on the quantification

of TGF-ß, TGF-ß assays were performed in the presence of DMEM-BSA containing rTGF-ß1 alone (closed squares), or with 0.5% (closed circles), 1% (closed triangles), or 2% (closed diamonds) calf serum. The rTGF-ß1 concentrations in the assays ranged from 0 to 8 pM. As shown in Figure 4C, serum similarly enhanced the induction of the PAI/L construct by rTGF-ß1 similar to that by purified growth factors as shown in Example 3C. At low rTGF-ß1 concentrations (<1 pM), addition of 0.5, 1 or 2% serum had little effect on the luciferase activity. As the rTGF-ß1 concentration was increased, the serum-containing curves were shifted upwards possibly as a result of growth factors such as bFGF in the serum.

E. Comparison of the TGF-ß Assay with the MLEC Assay and the Radioreceptor Assay for Quantifying TGF-ß

Quantification of TGF-ß in a defined media (DMEM-BSA) lacking growth factors or serum as demonstrated in Example 3D, however, is rarely found in the laboratory. For this reason, TGF-ß assays were also performed in COS, BSM and BAE cell conditioned medium (CM), all of which normally contain latent but little, if any, active TGF-ß. These samples were tested using the TGF-ß assay method of this invention in comparison with the MLEC (mink lung epithelial cell tritiated thymidine uptake cell assay).

The TGF-ß assay was performed as described in Example 3A with rTGF-ß1 ranging in concentration from 0 to 40 pM in the presence of either DMEM-BSA (closed squares), COS CM (crosses), BSM CM (closed triangles) or BAE CM (closed circles). To prepare conditioned medium, BAE cells were cultured in alphaMEM medium (Bio-Whittaker, Walkersville, MD) containing 5% fetal calf serum. BSM and COS cells were cultured in DMEM supplemented with 10% calf serum (Bio-Whittaker). Conditioned medium was prepared by a 24 hour incubation of the indicated cells with DMEM containing 0.1% pyrogen-poor BSA

(weight/volume) (Pierce, Rockford, IL). All media were supplemented with L-glutamine (2 mM), penicillin G (100 U/ml) and streptomycin sulfate (100 µg/ml) (Irvine Scientific, Santa Ana, CA).

The MLEC assay was performed essentially as described by Lucas et al., In Peptide Growth Factors, Barnes et al., Eds, Academic Press Inc. 198:303-316 (1991). Briefly, 100 ul aliguots of the samples were placed in 96-well plates containing 10^4 MLE cells per well in 100 ul of assay buffer (DMEM containing 0.25% fetal calf serum and 10 mM HEPES). After 20 hours at 37°C, one µCi of ^{3}H-thymidine (6.7Ci/mmol, Du Pont Co., Boston, MA) in 20 µl of the assay buffer was added to each well, and the plates incubated an additional 4 hours. The cells were harvested by incubation with 100 µl of 0.25% trypsin/1ml EDTA at 37°C for 15 minutes, transferred onto glass fiber filters, and placed into vials containing liquid scintillation solution. The amount of radioactivity was quantified with a Beckman LS 3801 ß-scintillation counter (Fullerton, CA).

As clearly shown by the data indicated by the unbroken lines in Figure 5, both BAE and BSM CM contained factors that stimulated thymidine incorporation in the MLEC assay 5-6 fold. Only at rTGF-ß1 levels greater than or equal to 1 pM was the ^{3}H-thymidine incorporation suppressed to a level equal to that of non-conditioned medium (DMEM-BSA). In contrast, COS CM contained factors that strongly inhibited ^{3}H-thymidine incorporation. With all three of these CM, calculation of TGF-ß concentration would be very difficult using ^{3}H-thymidine incorporation. In contrast, when different CM were used in the TGF-ß assay as indicated in Figure 5 with the data plotted with broken lines, there were also slight changes but these differences were much less significant than those seen with the MLEC assay. BAE CM, which contains bFGF, shifted the response curve to higher values. BSM and COS CM had only minor effects on the standard curves.

When bFGF (closed diamonds), EGF (open circles), PDGF-BB (open triangles), rIL-1alpha (open squares), and the TGF-ßs (rTGF-ß1 (closed squares), rTGF-ß2 (closed circles), and rTGF-ß3 (closed triangles) were tested for their ability to affect ^{3}H-thymidine incorporation into non-transfected MLE cells in the MLEC assay performed as described above, more striking effects were observed as shown in Figure 6. The three TGF-ß isoforms, especially TGF-ß3, decreased ^{3}H-thymidine incorporation as expected. IL-1alpha and PDGF-BB had little effect, but bFGF and EGF had strong dose-dependent stimulatory effects on ^{3}H-thymidine incorporation. Such effects can make the MLEC assays inaccurate and difficult to analyze.

F. Quantitation of Total TGF-ß Levels in Activated

In order to analyze total levels of TGF-ß, BAE CM collected after 12 or 24 hours was heat treated at 80°C for 10-12 minutes to activate endogenous latent TGF-ß as described by Brown et al., Growth Fact., 3:35-43 (1990). After cooling, the samples were diluted to 5, 10 or 20% of their original concentration with DMEM-BSA and were quantified using the TGF-ß assay. TGF-ß concentrations of 23.4±3.4 pM (12 hour CM) and 122.1±16 pM (24 hours CM) were determined via comparison with a rTGF-ß standard reference curve generated from plotting the detected amounts of luciferase activity that resulted from a range of predetermined amounts of TGF-ß as described in Example 3A.

The heat-activated CM were also assayed using the highly specific radioreceptor assay as described by Kojima et al., J. Cell. Physiol., 155:323-332 (1993), the disclosure of which is hereby incorporated by reference. Briefly, murine AKR-2B fibroblasts at 1 X 10^5 cells/well were plated in a 24-well plate in McCoy's 5A medium (Gibco BRL) supplemented with 5% fetal calf serum. The following day, the cells were washed 3 times with binding buffer (McCoy's 5A, 0.1% BSA, 25 mM HEPES at pH 7.4) and were pre-incubated in 250 ul of binding buffer for

1 hour at room temperature. The medium was removed, and the cells were incubated for 2 hours at room temperature in a mixture of 125 ul of binding buffer containing 50 pM ^{125}I-rTGF-ß1 and an equal volume of heat-activated (80°C for 10 minutes) BAE CM or serial dilutions of cold rTGF-ß1. The cells were washed 3 times with binding buffer, and the bound radioactivity was solubilized in cell lysis buffer (Analytical Luminescence) and was measured in a Packard Multi-PRIAS1 gamma counter (Meriden, CT). The radioreceptor assay was sensitive between 0.0004 and 2 nM rTGF-ß1.

In the radioreceptor assay, concentrations of 24±1.1 pM (12 hour CM) and 128±48.8 pM (24 hour CM) were calculated. The essentially identical results quantifying the amount of TGF-ß in conditioned medium between the TGF-ß assay described above and the radioreceptor assay verify the accuracy and specificity of the TGF-ß assay of this invention.

Thus, a highly sensitive and specific, non-radioactive assay for mature TGF-ß has now been developed. When compared to the sensitive and widely used MLEC method for measuring TGF-ß concentration, the TGF-ß assay was more rapid, had comparable sensitivity, and a greater detection range. Specificity of this assay was also higher as evidenced by its relative insensitivity to factors such as EGF and bFGF which can greatly affect other assays. The most remarkable example of the TGF-ß assay specificity was observed with COS cell CM which completely inhibited the MLEC assay, while having no detrimental effects in the TGF-ß assay.

In addition to the TGF-ß assay of this invention and the MLEC and radioreceptor assays described herein, other assays have been used to detect mature TGF-ß including anchorage-independent growth assays, differentiation-based assays, cell migration and plasminogen activity assays, radioimmunoassays and enzyme-linked immunosorbent assays. Although all of these assays can detect mature TGF-ß, the low concentrations of TGF-ß, generally less than 2 pM, generated in many biological

systems make many of them impractical without prior concentration of the sample that can result in large losses of the mature growth factor or even activation of latent TGF-ß. The TGF-ß assay of this invention overcomes these deficiencies by being highly sensitive and specific as well as nonradioactive. The specificity and sensitivity of the assay are the result of using a truncated PAI-1 promoter beginning at -800 and extending through 76 of the PAI-1 5' promoter that retains two regions responsible for maximal response to TGF-ß as described by Keeton et al., J. Biol. Chem., 266:23048-23052 (1991). Use of the complete PAI-1 promoter and upstream elements result in decreased specificity as responsive elements for other molecules present in complex solutions may be activated or inhibited deleteriously effecting the ability to quantify TGF-ß. Moreover, the truncated PAI-1 promoter used above has been further fragmented to smaller more specific TGF-ß response elements as described in Example 4 to enhance specificity and increase the sensitivity of the TGF-ß assay method.

When the TGF-ß assay is compared to the sensitive and widely used MLEC assay for quantifying TGF-ß concentrations, the TGF-ß assay was more rapid, had comparable sensitivity but with a greater detection range. Specificity of the assay was also higher as evidenced by the TGF-ß's assay insensitivity to growth factors such as EGF and bFGF that have been shown to greatly effect other assays. The most striking example of the specificity of the TGF-ß assay was observed with the COS cell line conditioned medium that completely inhibited the MLEC assay while having no detrimental effects in the TGF-ß assay as shown in Figure 5.

Although the TGF-ß assay is not isoform specific, use of the appropriate standard reference curves and addition of neutralizing antibodies to the various TGF-ß species allows for quantification of unique isoforms. While the TGF-ß assay of this invention is highly specific, the use of highly specific

neutralizing antibodies to TGF-ß was used to verify that no other molecules were present in test liquid samples that may have affected the quantitation of TGF-ß in the assay. Considering its large range and specificity, this rapid, sensitive, non-radioactive, easily performed assay is of invaluable use in determining active TGF-ß concentrations in complex solutions, particularly so with the use of parallel assays with neutralizing antibodies to TGF-ß in complex unknown samples to verify that no other molecules are present that can affect the assay through either inhibition or activation of other regions of the truncated PAI-1 promoter.

4. Quantifying TGF-ß with Cells Transiently Transformed with Expression Vectors Having Shorter Fragments of the PAI-1 Promoter Containing TGF-ß Response Elements

The regulation of PAI-1 by TGF-ß appears to affect a number of biological systems and the mechanism of transcriptional regulation by TGF-ß has been studied by a number of groups. For example, the autoinduction of the TGF-ß1 promoter suggests a feedback loop designed to amplify the response to TGF-ß under certain conditions. This response was shown to involve specific AP-1 sites. AP-1 is a heterodimeric complex of Fos and Jun protein subunits that binds to specific DNA enhancer sites which have the consensus sequence TGASTCA (SEQ ID NO 26), where S can be either G or C. AP-1 is believed to mediate the transcriptional effects of the tumor promoting phorbol esters.

In contrast to these results, the TGF-ß response sequence in the promoter for type 1 collagen, has been localized to a sequence with homology to a nuclear factor 1 (NF-1) binding site. A number of different consensus sequences for NF-1 have been described and these include the sequences $TGGN_7GCCAA$ (SEQ ID NO 27), where N can be either A, C, G or T, and TGGCA (SEQ ID NO 28). The effect of TGF-ß on the PAI-1 promoter has been

studied resulting in the demonstration that the responsive regions contain sequences with homology to the AP-1 consensus sequence.

To determine the role of AP-1 in the regulation of the PAI-1 promoter in more detail and to identify smaller TGF-ß responsive regions with the PAI-1 promoter of p800neoLuc expression vector prepared in Example 1 for use in quantifying TGF-ß in Example 3, the effect of both TGF-ß and AP-1 on the activity of a 25 bp fragment corresponding to the PAI-1 promoter between -674 and -650 in the 5' flanking region was evaluated. This fragment contained one of the AP-1 like sequences that responded to TGF-ß. The expression vectors for use in assessing the requirement for AP-1, including the one containing the 25 bp fragment, were prepared as described in Example 1C.

A. TGF-ß Activation of PAI-1 Promoter Fragments

AP-1 like sites are located within each of three regions of the 5' flanking region of the PAI-1 promoter from -87 to -49, from -674 to -636 and from -740 to -703. Oligonucleotides having portions or all of these regions were synthesized and cloned into a pUC-luciferase expressing plasmid containing the minimal promoter as described in Example 1C. The resultant plasmids were transiently transfected into recipient Hep3B cells as described in Example 2C and evaluated for their response to TGF-ß as measured by luciferase expression as described in Example 3A. The plasmid designated p56Luc contained an oligonucleotide sequence that corresponded to -56 to -41 of the PAI-1 promoter gene (also referred to as region A) and conferred a 10-fold induction of measurable TGF-ß as compared to a 3-fold induction obtained with a plasmid expression vector only containing the minimal promoter sequence.

Another plasmid designated p674Luc, deposited with ATCC and having ATCC Accession Number 75627, contained an

oligonucleotide sequence 25 bp in length that corresponded to -674 to -650 of the PAI-1 promoter (also referred to as region B). This nucleotide sequence conferred a 70-fold induction on the minimal promoter. The plasmid designated p743Luc contained an oligonucleotide sequence 35 bp in length that corresponded to -743 to -708 of the PAI-1 promoter (also referred to as region C). This nucleotide sequence conferred a 35-fold induction in the promoter. The plasmid designated p732Luc exhibited 62-fold induction while the plasmid, p732HBV, having the hepatitis B virus (HBV) minimal promoter sequence instead of the PAI-1 sequence exhibited 47-fold induction.

This result is in comparison to 6-fold basal induction from a control plasmid having only the HBV minimal promoter without having any TGF-ß response elements. The nucleotide sequence of the sense strand of the HBV-minimal promoter-containing plasmid having or lacking the neomycin selectable marker gene are listed respectively in SEQ ID NOs 23 and 24. In parallel assays, the p800Luc plasmid that contained 3 AP-1-like sequences conferred greater than 150-fold induction of TGF-ß responsiveness as compared to the minimal promoter sequence. The stably transformed p1500Luc similarly resulted in approximately 150-fold induction. These results as well as the others presented in the Examples represent the average of at least 4 independent experiments, each performed in duplicate.

Regions A and C contained only a single AP-1 like sequence whereas region B contained 2 AP-1 like binding sequences. Thus, oligonucleotides containing AP-1 like sequences from each region were able to confer TGF-ß responsiveness to a non-responsive minimal promoter.

B. Responsiveness of the TGF-ß responsive Regions A, B and C to c-fos/c-jun

In order to directly test the response of the p56Luc, p674Luc and p743Luc plasmids to AP-1, they were cotransfected

together into Hep3B cells with plasmids containing the mouse genes for c-fos and c-jun under the control of the RSV promoter. All three of these regions showed a dose dependent response to increasing amounts of c-fos/c-jun, with maximum responses seen using 0.1 µg/well of c-fos and c-jun plasmids. This response was dependent on co-transfection of both plasmids since neither c-fos or c-jun alone was able to cause this induction.

C. Detailed Analysis of the TGF-ß Responsive Nucleotide Sequence in the PAI-1 Promoter from Nucleotide -743 to -708 (Region C)

To find the minimal TGF-ß responsive sequence in the PAI-1 promoter region from nucleotide position -743 to -708, the sequence of which is listed in SEQ ID NO 16, two oligonucleotides were made, the first from the 3' side of region C which contained the AP-1 like sequence (C2: -723 to -708 corresponding to the sequence in SEQ ID NO 16 from 21 to 36) and the second from the remaining 5' sequence (C3: -743 to -727 corresponding to the sequence in SEQ ID NO 16 from 1 to 17). When the oligonucleotides were examined for response to TGF-ß, neither the C2 or C3 sequence showed maximal induction with TGF-ß (10-fold and 3-fold induction, respectively) as compared to region C itself (25-fold induction). This result suggested that a portion of a TGF-ß responsive binding site located between -723 and -727 was deleted. The 5' side of C2 was then progressively extended to include bases between -723 to -728 (7-fold induction) but found that this did not improve the TGF-ß response. However when this region was extended another 4 bp there was a dramatic increase in the TGF-ß response (63-fold induction) indicating that this region was crucial to this response.

D. Site-Specific Mutations of the PAI-1 Promoter from Nucleotide -732 to -708, Region C5

To assess the role of the AP-1 site compared to the 5' TGF-ß responsive site, the response of the minimal promoter having the 5' flanking region of the PAI-1 promoter from -39 to +76 to direct stimulation with c-fos/c-jun was determined. It showed 10-fold induction with AP-1 compared to only 3-fold induction with TGF-ß. When C5 was tested in a similar manner there was only a 2-fold increase above the vector background induced by c-fos/c-jun compared to a greater than 20-fold increase above background seen with TGF-ß (C5 itself showed 63-fold induction). Thus, although the wild type AP-1 site in C5 was only a relatively poor responsive sequence to c-fos/c-jun, this region still showed a strong response to TGF-ß. The AP-1 site was therefore mutated to produce a consensus AP-1 sequence (TGACACA to TGAGTCA, SEQ ID NOs 29 and 30, respectively) and the response of mutant to both c-fos/c-jun and TGF-ß was compared. This mutation increased the AP-1 response from 19-fold to 105-fold but did not improve the TGF-ß response. In fact, a consistent decrease was seen in the TGF-ß response following this mutation (63-fold induction with TGF-ß for the wild type AP-1 like site to 30-fold for the consensus AP-1 site).

The AP-1 like site was then mutated by changing the critical TGA bases, a change shown by others to decrease the activity of the AP-1 binding site. Although this mutation had the expected effect of abolishing the AP-1 response, it did not completely abolish the response of this construct to TGF-ß (10-fold induction with c-fos/c-jun [i.e., vector background] but a 13-fold induction with TGF-ß [i.e., 5-fold above vector background]).

This result once again suggested that the 5' portion of C5 (-732 to -708) was more critical than the AP-1 like sequence in mediating the TGF-ß response. To further test this hypothesis, 4 bp between -728 and -732 was mutated (the resultant mutated vector designated C8) since the previous deletion results suggested that this sequence was critical to the TGF-ß

response. A 3 bp sequence between -726 and -728 was also mutated (the resultant vector was designated C9). As expected, both of these 5' mutations caused dramatic reductions in the response of C5 to TGF-ß (60-fold to 4-fold for both C8 and C9). These changes had little effect on the AP-1 response which decreased only slightly from 19-fold to 13-fold. A double mutation of both of these sites was also created and this abolished both the TGF-ß and the AP-1 activity.

E. Heterologous Promoter Induction

To test whether the 25 bp oligonucleotide from the PAI-1 promoter region C5, -732 to -708 (SEQ ID NO 15), was able to activate a heterologous promoter, it was cloned into a hepatitis B viral promoter, the latter of which had the nucleotide sequence from -188 to +145 of the viral promoter (SEQ ID NO 19). Control experiments found that this construct alone showed 28-fold induction with fos/jun. However, the viral promoter showed only 4-fold induction with TGF-ß. Thus, even though the hepatitis B viral promoter had active AP-1 like sites, these were not sufficient for a strong TGF-ß response.

The region between -708 and -732 of the PAI-1 promoter (C5) was then cloned into the viral promoter and the resultant construct was tested as above. The 25 bp PAI-1 fragment was able to dramatically increase the TGF-ß response of the viral promoter from 4-fold to 47-fold but did not alter the AP-1 response (25-fold compared to 28-fold). Finally, mutation of bases between -732 and -728 of the PAI-1 promoter oligonucleotide dramatically reduced the TGF-ß induction of this fragment but did not lower the response to AP-1.

F. AP-1-Independent TGF-ß Induction

To determine if the 5' -732 to -708 nucleotide sequence from the PAI-1 promoter could function independently of the AP-1 site in the TGF-ß response, a 15 bp oligonucleotide containing bases between -732 and -718, corresponding to the

nucleotide sequence from position 1 to 15 in SEQ ID NO 17) (which excludes the AP-1 like site) was cloned into a pUC-luciferase expression vector having the minimal PAI-1 promoter. This 15 bp sequence was able to confer 20-fold induction with TGF-ß with the minimal PAI-1 promoter and did not show any AP-1 activity.

With regard to the AP-1 like sites involved in this response, unlike the consensus sequence for AP-1 (TGASTCA, where S is G or C (SEQ ID NO 26), the most active sequences from the PAI-1 promoter all have the sequence TGA(N)ACA where N is either A, C, G or T (SEQ ID NO 31) (PAI-1 promoter: -717 to -711 = TGACACA (SEQ ID NO 29); -659 to -653 = TGATACA (SEQ ID NO 32). It is possible that the T to A substitution may affect the binding affinity enough to preferentially bind another protein other than c-fos/c-jun. This is consistent with the functional data on the AP-1 like site of the PAI-1 promoter (between -711 to -717) which indicates that the wild type sequence is a poor AP-1 binding site and yet is still important in the TGF-ß response.

The mutation and deletion data of the 25 bp sequence from the wild type PAI-1 promoter (-732 to -708) suggested that the 5' side of the oligonucleotide may contain a second binding site of importance in the TGF-ß response. In fact this region appeared to be more critical than the AP-1 sequence since mutation of this region almost completely abolished the TGF-ß response even though the AP-1 region was intact. When this sequence alone was evaluated, it was able to act independently of the AP-1 site and promote strong TGF-ß induction of the normally unresponsive minimal promoter. However, the full TGF-ß response was dependent on the functional activity of both the AP-1 like site and the 5' site. When the sequence of the 5' 15 bp sequence was compared to the other region of the PAI-1 promoter which also showed strong TGF-ß induction (region B = 60-fold), a sequence was found that was common to both of these regions (CCNTGTNT, where N is either A, C, G or T (SEQ ID NO

33)).

In summary, the TGF-ß response of the PAI-1 promoter has been localized to specific AP-1 like sites. However, the full TGF-ß response of this region of the PAI-1 promoter is dependent on the interaction of two binding sites. The first site has homology to an AP-1 site but does not appear to bind AP-1. While this site is not essential it is required for the full TGF-ß induction from this region. The second site, located 5' to the AP-1 site, appears to be critical in the TGF-ß response. This site is 15 bp in size and contains a motif that is present in both active regions of the PAI-1 promoter as well as in the most responsive region of the TGF-ß promoter. This novel sequence does not appear to match any previously described transcription factor binding sites and may represent a new and specific binding site which is critical for a strong TGF-ß response.

5. Deposit of Materials

The plasmids, p674Luc, p743Luc and p732Luc, were deposited on or before December 16, 1993, with the American Type Culture Collection, 1301 Parklawn Drive, Rockville, MD, USA (ATCC) and assigned the respective ATCC Accession Numbers ATCC 75627, ATCC 75628 and ATCC 75629. The cell line, Hep3B, stably transfected with plasmid p1500Luc for a transformed cell line designated LUCI, was also deposited on or before December 16, 1993 with ATCC and assigned the ATCC Accession Number CRL 11508. The deposit thus provides plasmids and a stably transfected cell line containing plasmid p1500Luc. These deposits were made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of viable plasmids and cell lines for 30 years from the date of deposit. The plasmids and cell line will be made available by ATCC under the terms of the Budapest Treaty which assures permanent and

unrestricted availability of the progeny of the culture to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 U.S.C. §122 and the Commissioner's rules pursuant thereto (including 37 CFR §1.14 with particular reference to 886 OG 638). The assignee of the present application has agreed that if the plasmid or cell line deposits should die or be lost or destroyed when cultivated under suitable conditions, they will be promptly replaced on notification with a viable specimen of the same plasmid or cell culture. Availability of the deposited plasmids is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the plasmids deposited, since the deposited embodiment is intended as a single illustration of one aspect of the invention and any plasmids that are functionally equivalent are within the scope of this invention. The deposit of material does not constitute an admission that the written description herein contained is inadequate to enable the practice of any aspect of the invention, including the best mode thereof, nor is it to be construed as limiting the scope of the claims to the specific illustration that it represents. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: The Scripps Research Institute
(B) STREET: 10666 North Torrey Pines Road
(C) CITY: La Jolla
(D) STATE: CA
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 92037
(G) TELEPHONE: 619-554-2937
(H) TELEFAX: 619-554-6312

(ii) TITLE OF INVENTION: A NEW SENSITIVE METHOD FOR QUANTIFYING ACTIVE TRANSFORMING GROWTH FACTOR-BETA AND COMPOSITIONS THEREFOR

(iii) NUMBER OF SEQUENCES: 33

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(v) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US 95/
(B) FILING DATE: 25-JAN-1995

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBERE: US 08/188,227
(B) FILING DATE: 25-JAN-1994

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11293 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATTCGA CCTCGAGGGA TCTTTGTGAA GGAACCTTAC TTCTGTGGTG 2160

TGACATAATT GGACAAACTA CCTACAGAGA TTTAAAGCTC TAAGGTAAAT ATAAAATTTT 2220

TAAGTGTATA ATGTGTTAAA CTACTGATTC TAATTGTTTG TGTATTTTAG ATTCCAACCT 2280

ATGGAACTGA TGAATGGGAG CAGTGGTGGA ATGCCTTTAA TGAGGAAAAC CTGTTTTGCT 2340

CAGAAGAAAT GCCATCTAGT GATGATGAGG CTACTGCTGA CTCTCAACAT TCTACTCCTC 2400

CAAAAAAGAA GAGAAAGGTA GAAGACCCCA AGGACTTTCC TTCAGAATTG CTAAGTTTTT 2460

TGAGTCATGC TGTGTTTAGT AATAGAACTC TTGCTTGCTT TGCTATTTAC ACCACAAAGG 2520

AAAAAGCTGC ACTGCTATAC AAGAAAATTA TGGAAAAATA TTCTGTAACC TTTATAAGTA 2580

GGCATAACAG TTATAATCAT AACATACTGT TTTTTCTTAC TCCACACAGG CATAGAGTGT 2640

CTGCTATTAA TAACTATGCT CAAAAATTGT GTACCTTTAG CTTTTTAATT TGTAAAGGGG 2700

TTAATAAGGA ATATTTGATG TATAGTGCCT TGACTAGAGA TCATAATCAG CCATACCACA 2760

TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC TCCCCCTGAA CCTGAAACAT 2820

AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG CTTATAATGG TTACAAATAA 2880

AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT 2940

TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC GGCTGTGGAA TGTGTGTCAG 3000

TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC 3060

AATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC CAGCAGGCAG AAGTATGCAA 3120

AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC 3180

CTAACTCCGC CCAGTTCCGC CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT 3240

GCAGAGGCCG AGGCCGCCTC GGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT 3300

GGAGGCCTAG GCTTTTGCAA AAAGCTTCAC GCTGCCGCAA GCACTCAGGG CGCAAGGGCT 3360

GCTAAAGGAA GCGGAACACG TAGAAAGCCA GTCCGCAGAA ACGGTGCTGA CCCCGGATGA 3420

ATGTCAGCTA CTGGGCTATC TGGACAAGGG AAAACGCAAG CGCAAAGAGA AAGCAGGTAG 3480

CTTGCAGTGG GCTTACATGG CGATAGCTAG ACTGGGCGGT TTTATGGACA GCAAGCGAAC 3540

CGGAATTGCC AGCTGGGGCG CCCTCTGGTA AGGTTGGGAA GCCCTGCAAA GTAAACTGGA 3600

TGGCTTTCTT GCCGCCAAGG ATCTGATGGC GCAGGGGATC AAGATCTGAT CAAGAGACAG 3660

GATGAGGATC GTTTCGCATG ATTGAACAAG ATGGATTGCA CGCAGGTTCT CCGGCCGCTT 3720

GGGTGGAGAG GCTATTCGGC TATGACTGGG CACAACAGAC AATCGGCTGC TCTGATGCCG 3780

CCGTGTTCCG GCTGTCAGCG CAGGGGCGCC CGGTTCTTTT TGTCAAGACC GACCTGTCCG 3840

GTGCCCTGAA TGAACTGCAG GACGAGGCAG CGCGGCTATC GTGGCTGGCC ACGACGGGCG 3900

TTCCTTGCGC AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG 3960

GCGAAGTGCC GGGGCAGGAT CTCCTGTCAT CTCACCTTGC TCCTGCCGAG AAAGTATCCA 4020

TCATGGCTGA TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC 4080

ACCAAGCGAA ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC 4140

AGGATGATCT GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTC GCCAGGCTCA 4200

AGGCGCGCAT GCCCGACGGC GAGGATCTCG TCGTGACCCA TGGCGATGCC TGCTTGCCGA 4260

ATATCATGGT GGAAAATGGC CGCTTTTCTG GATTCATCGA CTGTGGCCGG CTGGGTGTGG 4320

CGGACCGCTA TCAGGACATA GCGTTGGCTA CCCGTGATAT TGCTGAAGAG CTTGGCGGCG 4380

AATGGGCTGA CCGCTTCCTC GTGCTTTACG GTATCGCCGC TCCCGATTCG CAGCGCATCG 4440

CCTTCTATCG CCTTCTTGAC GAGTTCTTCT GAGCGGGACT CTGGGGTTCG AAATGACCGA 4500

CCAAGCGACG CCCAACCTGC CATCACGAGA TTTCGATTCC ACCGCCGCCT TCTATGAAAG 4560

GTTGGGCTTC GGAATCGTTT TCCGGGACGC CGGCTGGATG ATCCTCCAGC GCGGGGATCT 4620

CATGCTGGAG TTCTTCGCCC ACCCCGGGCT CGATCCCCTC GCGAGTTGGT TCAGCTGCTG 4680

CCTGAGGCTG GACGACCTCG CGGAGTTCTA CCGGCAGTGC AAATCCGTCG GCATCCAGGA 4740

AACCAGCAGC GGCTATCCGC GCATCCATGC CCCCGAACTG CAGGAGTGGG GAGGCACGAT 4800

GGCCGCTTTG GTCCCGGATC TTTGTGAAGG AACCTTACTT CTGTGGTGTG ACATAATTGG 4860

ACAAACTACC TACAGAGATT TAAAGCTCTA AGGTAAATAT AAAATTTTTA AGTGTATAAT 4920

GTGTTAAACT ACTGATTCTA ATTGTTTGTG TATTTTAGAT TCCAACCTAT GGAACTGATG 4980

AATGGGAGCA GTGGTGGAAT GCCTTTAATG AGGAAAACCT GTTTTGCTCA GAAGAAATGC 5040

CATCTAGTGA TGATGAGGCT ACTGCTGACT CTCAACATTC TACTCCTCCA AAAAAGAAGA 5100

GAAAGGTAGA AGACCCCAAG GACTTTCCTT CAGAATTGCT AAGTTTTTTG AGTCATGCTG 5160

TGTTTAGTAA TAGAACTCTT GCTTGCTTTG CTATTTACAC CACAAAGGAA AAAGCTGCAC 5220

TGCTATACAA GAAAATTATG GAAAAATATT CTGTAACCTT TATAAGTAGG CATAACAGTT 5280

ATAATCATAA CATACTGTTT TTTCTTACTC CACACAGGCA TAGAGTGTCT GCTATTAATA 5340

ACTATGCTCA AAAATTGTGT ACCTTTAGCT TTTTAATTTG TAAAGGGGTT AATAAGGAAT 5400

ATTTGATGTA TAGTGCCTTG ACTAGAGATC ATAATCAGCC ATACCACATT TGTAGAGGTT 5460

TTACTTGCTT TAAAAAACCT CCCACACCTC CCCCTGAACC TGAAACATAA AATGAATGCA 5520

ATTGTTGTTG TTAACTTGTT TATTGCAGCT TATAATGGTT ACAAATAAAG CAATAGCATC 5580

ACAAATTTCA CAAATAAAGC ATTTTTTTCA CTGCATTCTA GTTGTGGTTT GTCCAAACTC 5640

ATCAATGTAT CTTATCATGT CTGGATCCCC AGGAAGCTCC TCTGTGTCCT CATAAACCCT 5700

AACCTCCTCT ACTTGAGAGG ACATTCCAAT CATAGGCTGC CCATCCACCC TCTGTGTCCT 5760

CCTGTTAATT AGGTCACTTA ACAAAAAGGA AATTGGGTAG GGGTTTTTCA CAGACCGCTT 5820

TCTAAGGGTA ATTTTAAAAT ATCTGGGAAG TCCCTTCCAC TGCTGTGTTC CAGAAGTGTT 5880

GGTAAACAGC CCACAAATGT CAACAGCAGA AACATACAAG CTGTCAGCTT TGCACAAGGG 5940

CCCAACACCC TGCTCATCAA GAAGCACTGT GGTTGCTGTG TTAGTAATGT GCAAAACAGG 6000

AGGCACATTT TCCCCACCTG TGTAGGTTCC AAAATATCTA GTGTTTTCAT TTTTACTTGG 6060

ATCAGGAACC CAGCACTCCA CTGGATAAGC ATTATCCTTA TCCAAAACAG CCTTGTGGTC 6120

AGTGTTCATC TGCTGACTGT CAACTGTAGC ATTTTTTGGG GTTACAGTTT GAGCAGGATA 6180

TTTGGTCCTG TAGTTTGCTA ACACACCCTG CAGCTCCAAA GGTTCCCCAC CAACAGCAAA 6240

AAAATGAAAA TTTGACCCTT GAATGGGTTT TCCAGCACCA TTTTCATGAG TTTTTTGTGT 6300

CCCTGAATGC AAGTTTAACA TAGCAGTTAC CCCAATAACC TCAGTTTTAA CAGTAACAGC 6360

TTCCCACATC AAAATATTTC CACAGGTTAA GTCCTCATTT AAATTAGGCA AAGGAATTAT 6420

ACACTCCGCT ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG 6480

CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG 6540

TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGGCAGC 6600

GGATCATAAT CAGCCATACC ACATTTGTAG AGGTTTTACT TGCTTTAAAA AACCTCCCAC 6660

ACCTCCCCCT GAACCTGAAA CATAAAATGA ATGCAATTGT TGTTGTTAAC TTGTTTATTG 6720

CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT 6780

TTTCACTGCA TTCTAGTTGT GGTTTGTCCA AACTCATCAA TGTATCTTAT CATGTCTGGA 6840

TCATAATCAG CCATACCACA TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC 6900

TCCCCCTGAA CCTGAAACAT AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG 6960

CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT 7020

CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC 7080

CAAGCTTACC ATGGTAACCC CTGGTCCCGT TCAGCCACCA CCACCCCACC CAGCACACCT 7140

CCAACCTCAG CCAGACAAGG TTGTTGACAC AAGAGAGCCC TCAGGGGCAC AGAGAGAGTC 7200

TGGACACGTG GGGAGTCAGC CGTGTATCAT CGGAGGCGGC CGGGCACATG GCAGGGATGA 7260

GGGAAAGACC AAGAGTCCTC TGTTGGGCCC AAGTCCTAGA CAGACAAAAC CTAGACAATC 7320

ACGTGGCTGG CTGCATGCCT GTGGCTGTTG GGCTGGGCAG GAGGAGGGAG GGGCGCTCTT 7380

TCCTGGAGGT GGTCCAGAGC ACCGGGTGGA CAGCCCTGGG GGAAAACTTC CACGTTTTGA 7440

TGGAGGTTAT CTTTGATAAC TCCACAGTGA CCTGGTTCGC CAAAGGAAAA GCAGGCAACG 7500

TGAGCTGTTT TTTTTTTCTC CAAGCTGAAC ACTAGGGGTC CTAGGCTTTT TGGGTCACCC 7560

GGCATGGCAG ACAGTCAACC TGGCAGGACA TCCGGGAGAG ACAGACACAG GCAGAGGGCA 7620

GAAAGGTCAA GGGAGGTTCT CAGGCCAAGG CTATTGGGGT TTGCTCAATT GTTCCTGAAT 7680

GCTCTTACAC ACGTACACAC ACAGAGCAGC ACACACACAC ACACACACAT GCCTCAGCAA 7740

GTCCCAGAGA GGGAGGTGTC GAGGGGGACC CGCTGGCTGT TCAGACGGAC TCCCAGAGCC 7800

AGTGAGTGGG TGGGGCTGGA ACATGAGTTC ATCTATTTCC TGCCCACATC TGGTATAAAA 7860

GGAGGCAGTG GCCCACAGAG GAGCACAGCT GTGTTTGGCT GCAGGGCCAA GAGCGCTGTC 7920

AAGAAGACCC ACACGCCCCC CTCCAGCAGC TGAATTCCAG CTGGCATTCC GGTACTGTTG 7980

GTAAAATGGA AGACGCCAAA AACATAAAGA AAGGCCCGGC GCCATTCTAT CCTCTAGAGG 8040

ATGGAACCGC TGGAGAGCAA CTGCATAAGG CTATGAAGAG ATACGCCCTG GTTCCTGGAA 8100

CAATTGCTTT TACAGATGCA CATATCGAGG TGAACATCAC GTACGCGGAA TACTTCGAAA 8160

TGTCCGTTCG GTTGGCAGAA GCTATGAAAC GATATGGGCT GAATACAAAT CACAGAATCG 8220

TCGTATGCAG TGAAAACTCT CTTCAATTCT TTATGCCGGT GTTGGGCGCG TTATTTATCG 8280

GAGTTGCAGT TGCGCCCGCG AACGACATTT ATAATGAACG TGAATTGCTC AACAGTATGA 8340

ACATTTCGCA GCCTACCGTA GTGTTTGTTT CCAAAAAGGG GTTGCAAAAA ATTTTGAACG 8400

TGCAAAAAAA ATTACCAATA ATCCAGAAAA TTATTATCAT GGATTCTAAA ACGGATTACC 8460

AGGGATTTCA GTCGATGTAC ACGTTCGTCA CATCTCATCT ACCTCCCGGT TTTAATGAAT 8520

ACGATTTTGT ACCAGAGTCC TTTGATCGTG ACAAAACAAT TGCACTGATA ATGAATTCCT 8580

CTGGATCTAC TGGGTTACCT AAGGGTGTGG CCCTTCCGCA TAGAACTGCC TGCGTCAGAT 8640

TCTCGCATGC CAGAGATCCT ATTTTTGGCA ATCAAATCAT TCCGGATACT GCGATTTTAA 8700

GTGTTGTTCC ATTCCATCAC GGTTTTGGAA TGTTTACTAC ACTCGGATAT TTGATATGTG 8760

GATTTCGAGT CGTCTTAATG TATAGATTTG AAGAAGAGCT GTTTTTACGA TCCCTTCAGG 8820

ATTACAAAAT TCAAAGTGCG TTGCTAGTAC CAACCCTATT TTCATTCTTC GCCAAAAGCA 8880

CTCTGATTGA CAAATACGAT TTATCTAATT TACACGAAAT TGCTTCTGGG GGCGCACCTC 8940

TTTCGAAAGA AGTCGGGGAA GCGGTTGCAA AACGCTTCCA TCTTCCAGGG ATACGACAAG 9000

GATATGGGCT CACTGAGACT ACATCAGCTA TTCTGATTAC ACCCGAGGGG GATGATAAAC 9060

CGGGCGCGGT CGGTAAAGTT GTTCCATTTT TTGAAGCGAA GGTTGTGGAT CTGGATACCG 9120

GGAAAACGCT GGGCGTTAAT CAGAGAGGCG AATTATGTGT CAGAGGACCT ATGATTATGT 9180

CCGGTTATGT AAACAATCCG GAAGCGACCA ACGCCTTGAT TGACAAGGAT GGATGGCTAC 9240

ATTCTGGAGA CATAGCTTAC TGGGACGAAG ACGAACACTT CTTCATAGTT GACCGCTTGA 9300

AGTCTTTAAT TAAATACAAA GGATATCAGG TGGCCCCCGC TGAATTGGAA TCGATATTGT 9360

TACAACACCC CAACATCTTC GACGCGGGCG TGGCAGGTCT TCCCGACGAT GACGCCGGTG 9420

AACTTCCCGC CGCCGTTGTT GTTTTGGAGC ACGGAAAGAC GATGACGGAA AAAGAGATCG 9480

TGGATTACGT CGCCAGTCAA GTAACAACCG CGAAAAAGTT GCGCGGAGGA GTTGTGTTTG 9540

TGGACGAAGT ACCGAAAGGT CTTACCGGAA AACTCGACGC AAGAAAAATC AGAGAGATCC 9600

TCATAAAGGC CAAGAAGGGC GGAAAGTCCA AATTGTAAAA TGTAACTGTA TTCAGCGATG 9660

ACGAAATTCT TAGCTATTGT AATGACTCTA GAGGATCTTT GTGAAGGAAC CTTACTTCTG 9720

TGGTGTGACA TAATTGGACA AACTACCTAC AGAGATTTAA AGCTCTAAGG TAAATATAAA 9780

ATTTTTAAGT GTATAATGTG TTAAACTACT GATTCTAATT GTTTGTGTAT TTTAGATTCC 9840

AACCTATGGA ACTGATGAAT GGGAGCAGTG GTGGAATGCC TTTAATGAGG AAAACCTGTT 9900

TTGCTCAGAA GAAATGCCAT CTAGTGATGA TGAGGCTACT GCTGACTCTC AACATTCTAC 9960

TCCTCCAAAA AAGAAGAGAA AGGTAGAAGA CCCCAAGGAC TTTCCTTCAG AATTGCTAAG 10020

TTTTTTGAGT CATGCTGTGT TTAGTAATAG AACTCTTGCT TGCTTTGCTA TTTACACCAC 10080

AAAGGAAAAA GCTGCACTGC TATACAAGAA AATTATGGAA AAATATTCTG TAACCTTTAT 10140

AAGTAGGCAT AACAGTTATA ATCATAACAT ACTGTTTTTT CTTACTCCAC ACAGGCATAG 10200

AGTGTCTGCT ATTAATAACT ATGCTCAAAA ATTGTGTACC TTTAGCTTTT TAATTTGTAA 10260

AGGGGTTAAT AAGGAATATT TGATGTATAG TGCCTTGACT AGAGATCATA ATCAGCCATA 10320

CCACATTTGT AGAGGTTTTA CTTGCTTTAA AAAACCTCCC ACACCTCCCC CTGAACCTGA 10380

AACATAAAAT GAATGCAATT GTTGTTGTTA ACTTGTTTAT TGCAGCTTAT AATGGTTACA 10440

AATAAAGCAA TAGCATCACA AATTTCACAA ATAAAGCATT TTTTTCACTG CATTCTAGTT 10500

GTGGTTTGTC CAAACTCATC AATGTATCTT ATCATGTCTG GATCCCCAGG AAGCTCCTCT 10560

GTGTCCTCAT AAACCCTAAC CTCCTCTACT TGAGAGGACA TTCCAATCAT AGGCTGCCCA 10620

TCCACCCTCT GTGTCCTCCT GTTAATTAGG TCACTTAACA AAAAGGAAAT TGGGTAGGGG 10680

TTTTTCACAG ACCGCTTTCT AAGGGTAATT TTAAAATATC TGGGAAGTCC CTTCCACTGC 10740

TGTGTTCCAG AAGTGTTGGT AAACAGCCCA CAAATGTCAA CAGCAGAAAC ATACAAGCTG 10800

TCAGCTTTGC ACAAGGGCCC AACACCCTGC TCAGCAAGAA GCACTGTGGT TGCTGTGTTA 10860

GTAATGTGCA AAACAGGAGG CACATTTTCC CCACCTGTGT AGGTTCCAAA ATATCTAGTG 10920

TTTTCATTTT TACTTGGATC AGGAACCCAG CACTCCACTG GATAAGCATT ATCCTTATCC 10980

AAAACAGCCT TGTGGTCAGT GTTCATCTGC TGACTGTCAA CTGTAGCATT TTTTGGGGTT 11040

ACAGTTTGAG CAGGATATTT GGTCCTGTAG TTTGCTAACA CACCCTGCAG CTCCAAAGGT 11100

TCCCCACCAA CAGCAAAAAA ATGAAAATTT GACCCTTGAA TGGGTTTTCC AGCACCATTT 11160

TCATGAGTTT TTTGTGTCCC TGAATGCAAG TTTAACATAG CAGTTACCCC AATAACCTCA 11220

GTTTTAACAG TAACAGCTTC CCACATCAAA ATATTTCCAC AGGTTAAGTC CTCATTTAAA 11280

TTAGGCAAAG GAA 11293

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10697 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATTCGA CCTCGAGGGA TCTTTGTGAA GGAACCTTAC TTCTGTGGTG 2160

TGACATAATT GGACAAACTA CCTACAGAGA TTTAAAGCTC TAAGGTAAAT ATAAAATTTT 2220

TAAGTGTATA ATGTGTTAAA CTACTGATTC TAATTGTTTG TGTATTTTAG ATTCCAACCT 2280

ATGGAACTGA TGAATGGGAG CAGTGGTGGA ATGCCTTTAA TGAGGAAAAC CTGTTTTGCT 2340

CAGAAGAAAT GCCATCTAGT GATGATGAGG CTACTGCTGA CTCTCAACAT TCTACTCCTC 2400

CAAAAAAGAA GAGAAAGGTA GAAGACCCCA AGGACTTTCC TTCAGAATTG CTAAGTTTTT 2460

TGAGTCATGC TGTGTTTAGT AATAGAACTC TTGCTTGCTT TGCTATTTAC ACCACAAAGG 2520

AAAAAGCTGC ACTGCTATAC AAGAAAATTA TGGAAAAATA TTCTGTAACC TTTATAAGTA 2580

GGCATAACAG TTATAATCAT AACATACTGT TTTTTCTTAC TCCACACAGG CATAGAGTGT 2640

CTGCTATTAA TAACTATGCT CAAAAATTGT GTACCTTTAG CTTTTTAATT TGTAAAGGGG 2700

TTAATAAGGA ATATTTGATG TATAGTGCCT TGACTAGAGA TCATAATCAG CCATACCACA 2760

TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC TCCCCCTGAA CCTGAAACAT 2820

AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG CTTATAATGG TTACAAATAA 2880

AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT 2940

TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC GGCTGTGGAA TGTGTGTCAG 3000

TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC 3060

AATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC CAGCAGGCAG AAGTATGCAA 3120

AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC 3180

CTAACTCCGC CCAGTTCCGC CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT 3240

GCAGAGGCCG AGGCCGCCTC GGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT 3300

GGAGGCCTAG GCTTTTGCAA AAAGCTTCAC GCTGCCGCAA GCACTCAGGG CGCAAGGGCT 3360

GCTAAAGGAA GCGGAACACG TAGAAAGCCA GTCCGCAGAA ACGGTGCTGA CCCCGGATGA 3420

ATGTCAGCTA CTGGGCTATC TGGACAAGGG AAAACGCAAG CGCAAAGAGA AAGCAGGTAG 3480

CTTGCAGTGG GCTTACATGG CGATAGCTAG ACTGGGCGGT TTTATGGACA GCAAGCGAAC 3540

CGGAATTGCC AGCTGGGGCG CCCTCTGGTA AGGTTGGGAA GCCCTGCAAA GTAAACTGGA 3600

TGGCTTTCTT GCCGCCAAGG ATCTGATGGC GCAGGGGATC AAGATCTGAT CAAGAGACAG 3660

GATGAGGATC GTTTCGCATG ATTGAACAAG ATGGATTGCA CGCAGGTTCT CCGGCCGCTT 3720

GGGTGGAGAG GCTATTCGGC TATGACTGGG CACAACAGAC AATCGGCTGC TCTGATGCCG 3780

CCGTGTTCCG GCTGTCAGCG CAGGGGCGCC CGGTTCTTTT TGTCAAGACC GACCTGTCCG 3840

GTGCCCTGAA TGAACTGCAG GACGAGGCAG CGCGGCTATC GTGGCTGGCC ACGACGGGCG 3900

TTCCTTGCGC AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG 3960

GCGAAGTGCC GGGGCAGGAT CTCCTGTCAT CTCACCTTGC TCCTGCCGAG AAAGTATCCA 4020

TCATGGCTGA TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC 4080

ACCAAGCGAA ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC 4140

AGGATGATCT GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTC GCCAGGCTCA 4200

AGGCGCGCAT GCCCGACGGC GAGGATCTCG TCGTGACCCA TGGCGATGCC TGCTTGCCGA 4260

ATATCATGGT GGAAAATGGC CGCTTTTCTG GATTCATCGA CTGTGGCCGG CTGGGTGTGG 4320

CGGACCGCTA TCAGGACATA GCGTTGGCTA CCCGTGATAT TGCTGAAGAG CTTGGCGGCG 4380

AATGGGCTGA CCGCTTCCTC GTGCTTTACG GTATCGCCGC TCCCGATTCG CAGCGCATCG 4440

CCTTCTATCG CCTTCTTGAC GAGTTCTTCT GAGCGGGACT CTGGGGTTCG AAATGACCGA 4500

CCAAGCGACG CCCAACCTGC CATCACGAGA TTTCGATTCC ACCGCCGCCT TCTATGAAAG 4560

GTTGGGCTTC GGAATCGTTT TCCGGGACGC CGGCTGGATG ATCCTCCAGC GCGGGGATCT 4620

CATGCTGGAG TTCTTCGCCC ACCCCGGGCT CGATCCCCTC GCGAGTTGGT TCAGCTGCTG 4680

CCTGAGGCTG GACGACCTCG CGGAGTTCTA CCGGCAGTGC AAATCCGTCG GCATCCAGGA 4740

AACCAGCAGC GGCTATCCGC GCATCCATGC CCCCGAACTG CAGGAGTGGG GAGGCACGAT 4800

GGCCGCTTTG GTCCCGGATC TTTGTGAAGG AACCTTACTT CTGTGGTGTG ACATAATTGG 4860

ACAAACTACC TACAGAGATT TAAAGCTCTA AGGTAAATAT AAAATTTTTA AGTGTATAAT 4920

GTGTTAAACT ACTGATTCTA ATTGTTTGTG TATTTTAGAT TCCAACCTAT GGAACTGATG 4980

AATGGGAGCA GTGGTGGAAT GCCTTTAATG AGGAAAACCT GTTTTGCTCA GAAGAAATGC 5040

CATCTAGTGA TGATGAGGCT ACTGCTGACT CTCAACATTC TACTCCTCCA AAAAAGAAGA 5100

GAAAGGTAGA AGACCCCAAG GACTTTCCTT CAGAATTGCT AAGTTTTTTG AGTCATGCTG 5160

TGTTTAGTAA TAGAACTCTT GCTTGCTTTG CTATTTACAC CACAAAGGAA AAAGCTGCAC 5220

TGCTATACAA GAAAATTATG GAAAAATATT CTGTAACCTT TATAAGTAGG CATAACAGTT 5280

ATAATCATAA CATACTGTTT TTTCTTACTC CACACAGGCA TAGAGTGTCT GCTATTAATA 5340

ACTATGCTCA AAAATTGTGT ACCTTTAGCT TTTTAATTTG TAAAGGGGTT AATAAGGAAT 5400

ATTTGATGTA TAGTGCCTTG ACTAGAGATC ATAATCAGCC ATACCACATT TGTAGAGGTT 5460

TTACTTGCTT TAAAAAACCT CCCACACCTC CCCCTGAACC TGAAACATAA AATGAATGCA 5520

ATTGTTGTTG TTAACTTGTT TATTGCAGCT TATAATGGTT ACAAATAAAG CAATAGCATC 5580

ACAAATTTCA CAAATAAAGC ATTTTTTTCA CTGCATTCTA GTTGTGGTTT GTCCAAACTC 5640

ATCAATGTAT CTTATCATGT CTGGATCCCC AGGAAGCTCC TCTGTGTCCT CATAAACCCT 5700

AACCTCCTCT ACTTGAGAGG ACATTCCAAT CATAGGCTGC CCATCCACCC TCTGTGTCCT 5760

CCTGTTAATT AGGTCACTTA ACAAAAAGGA AATTGGGTAG GGGTTTTTCA CAGACCGCTT 5820

TCTAAGGGTA ATTTTAAAAT ATCTGGGAAG TCCCTTCCAC TGCTGTGTTC CAGAAGTGTT 5880

GGTAAACAGC CCACAAATGT CAACAGCAGA AACATACAAG CTGTCAGCTT TGCACAAGGG 5940

CCCAACACCC TGCTCATCAA GAAGCACTGT GGTTGCTGTG TTAGTAATGT GCAAAACAGG 6000

AGGCACATTT TCCCCACCTG TGTAGGTTCC AAAATATCTA GTGTTTTCAT TTTTACTTGG 6060

ATCAGGAACC CAGCACTCCA CTGGATAAGC ATTATCCTTA TCCAAAACAG CCTTGTGGTC 6120

AGTGTTCATC TGCTGACTGT CAACTGTAGC ATTTTTTGGG GTTACAGTTT GAGCAGGATA 6180

TTTGGTCCTG TAGTTTGCTA ACACACCCTG CAGCTCCAAA GGTTCCCCAC CAACAGCAAA 6240

AAAATGAAAA TTTGACCCTT GAATGGGTTT TCCAGCACCA TTTTCATGAG TTTTTTGTGT 6300

CCCTGAATGC AAGTTTAACA TAGCAGTTAC CCCAATAACC TCAGTTTTAA CAGTAACAGC 6360

TTCCCACATC AAAATATTTC CACAGGTTAA GTCCTCATTT AAATTAGGCA AAGGAATTAT 6420

ACACTCCGCT ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG 6480

CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG 6540

TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGGCAGC 6600

GGATCATAAT CAGCCATACC ACATTTGTAG AGGTTTTACT TGCTTTAAAA AACCTCCCAC 6660

ACCTCCCCCT GAACCTGAAA CATAAAATGA ATGCAATTGT TGTTGTTAAC TTGTTTATTG 6720

CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT 6780

TTTCACTGCA TTCTAGTTGT GGTTTGTCCA AACTCATCAA TGTATCTTAT CATGTCTGGA 6840

TCATAATCAG CCATACCACA TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC 6900

TCCCCCTGAA CCTGAAACAT AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG 6960

CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT 7020

CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC 7080

CAAGCTTACC ATGGTAACCC CTGGTCCCGT TCAGCCACCA CCACCCCACC CAGCACACCT 7140

CCAACCTCAG CCAGACAAGG TTGTTGACAC AAGAGAGCCC TCAGGGGCAC AGAGAGAGTC 7200

TGGACACGTG GGGAGTCAGC CGTGTATCAT CGGAGGCGGC CGGGCACCCA CATCTGGTAT 7260

AAAAGGAGGC AGTGGCCCAC AGAGGAGCAC AGCTGTGTTT GGCTGCAGGG CCAAGAGCGC 7320

TGTCAAGAAG ACCCACACGC CCCCCTCCAG CAGCTGAATT CCAGCTGGCA TTCCGGTACT 7380

GTTGGTAAAA TGGAAGACGC CAAAAACATA AAGAAAGGCC CGGCGCCATT CTATCCTCTA 7440

GAGGATGGAA CCGCTGGAGA GCAACTGCAT AAGGCTATGA AGAGATACGC CCTGGTTCCT 7500

GGAACAATTG CTTTTACAGA TGCACATATC GAGGTGAACA TCACGTACGC GGAATACTTC 7560

GAAATGTCCG TTCGGTTGGC AGAAGCTATG AAACGATATG GGCTGAATAC AAATCACAGA 7620

ATCGTCGTAT GCAGTGAAAA CTCTCTTCAA TTCTTTATGC CGGTGTTGGG CGCGTTATTT 7680

ATCGGAGTTG CAGTTGCGCC CGCGAACGAC ATTTATAATG AACGTGAATT GCTCAACAGT 7740

ATGAACATTT CGCAGCCTAC CGTAGTGTTT GTTTCCAAAA AGGGGTTGCA AAAAATTTTG 7800

AACGTGCAAA AAAAATTACC AATAATCCAG AAAATTATTA TCATGGATTC TAAAACGGAT 7860

TACCAGGGAT TTCAGTCGAT GTACACGTTC GTCACATCTC ATCTACCTCC CGGTTTTAAT 7920

GAATACGATT TTGTACCAGA GTCCTTTGAT CGTGACAAAA CAATTGCACT GATAATGAAT 7980

TCCTCTGGAT CTACTGGGTT ACCTAAGGGT GTGGCCCTTC CGCATAGAAC TGCCTGCGTC 8040

AGATTCTCGC ATGCCAGAGA TCCTATTTTT GGCAATCAAA TCATTCCGGA TACTGCGATT 8100

TTAAGTGTTG TTCCATTCCA TCACGGTTTT GGAATGTTTA CTACACTCGG ATATTTGATA 8160

TGTGGATTTC GAGTCGTCTT AATGTATAGA TTTGAAGAAG AGCTGTTTTT ACGATCCCTT 8220

CAGGATTACA AAATTCAAAG TGCGTTGCTA GTACCAACCC TATTTTCATT CTTCGCCAAA 8280

AGCACTCTGA TTGACAAATA CGATTTATCT AATTTACACG AAATTGCTTC TGGGGGCGCA 8340

CCTCTTTCGA AAGAAGTCGG GGAAGCGGTT GCAAAACGCT TCCATCTTCC AGGGATACGA 8400

CAAGGATATG GGCTCACTGA GACTACATCA GCTATTCTGA TTACACCCGA GGGGGATGAT 8460

AAACCGGGCG CGGTCGGTAA AGTTGTTCCA TTTTTTGAAG CGAAGGTTGT GGATCTGGAT 8520

ACCGGGAAAA CGCTGGGCGT TAATCAGAGA GGCGAATTAT GTGTCAGAGG ACCTATGATT 8580

ATGTCCGGTT ATGTAAACAA TCCGGAAGCG ACCAACGCCT TGATTGACAA GGATGGATGG 8640

CTACATTCTG GAGACATAGC TTACTGGGAC GAAGACGAAC ACTTCTTCAT AGTTGACCGC 8700

TTGAAGTCTT TAATTAAATA CAAAGGATAT CAGGTGGCCC CCGCTGAATT GGAATCGATA 8760

TTGTTACAAC ACCCCAACAT CTTCGACGCG GGCGTGGCAG GTCTTCCCGA CGATGACGCC 8820

GGTGAACTTC CCGCCGCCGT TGTTGTTTTG GAGCACGGAA AGACGATGAC GGAAAAAGAG 8880

ATCGTGGATT ACGTCGCCAG TCAAGTAACA ACCGCGAAAA AGTTGCGCGG AGGAGTTGTG 8940

TTTGTGGACG AAGTACCGAA AGGTCTTACC GGAAAACTCG ACGCAAGAAA AATCAGAGAG 9000

ATCCTCATAA AGGCCAAGAA GGGCGGAAAG TCCAAATTGT AAAATGTAAC TGTATTCAGC 9060

GATGACGAAA TTCTTAGCTA TTGTAATGAC TCTAGAGGAT CTTTGTGAAG GAACCTTACT 9120

TCTGTGGTGT GACATAATTG GACAAACTAC CTACAGAGAT TTAAAGCTCT AAGGTAAATA 9180

TAAAATTTTT AAGTGTATAA TGTGTTAAAC TACTGATTCT AATTGTTTGT GTATTTTAGA 9240

TTCCAACCTA TGGAACTGAT GAATGGGAGC AGTGGTGGAA TGCCTTTAAT GAGGAAAACC 9300

TGTTTTGCTC AGAAGAAATG CCATCTAGTG ATGATGAGGC TACTGCTGAC TCTCAACATT 9360

CTACTCCTCC AAAAAAGAAG AGAAAGGTAG AAGACCCCAA GGACTTTCCT TCAGAATTGC 9420

TAAGTTTTTT GAGTCATGCT GTGTTTAGTA ATAGAACTCT TGCTTGCTTT GCTATTTACA 9480

CCACAAAGGA AAAAGCTGCA CTGCTATACA AGAAAATTAT GGAAAAATAT TCTGTAACCT 9540

TTATAAGTAG GCATAACAGT TATAATCATA ACATACTGTT TTTTCTTACT CCACACAGGC 9600

ATAGAGTGTC TGCTATTAAT AACTATGCTC AAAAATTGTG TACCTTTAGC TTTTTAATTT 9660

GTAAAGGGGT TAATAAGGAA TATTTGATGT ATAGTGCCTT GACTAGAGAT CATAATCAGC 9720

CATACCACAT TTGTAGAGGT TTTACTTGCT TTAAAAAACC TCCCACACCT CCCCCTGAAC 9780

CTGAAACATA AAATGAATGC AATTGTTGTT GTTAACTTGT TTATTGCAGC TTATAATGGT 9840

TACAAATAAA GCAATAGCAT CACAAATTTC ACAAATAAAG CATTTTTTTC ACTGCATTCT 9900

AGTTGTGGTT TGTCCAAACT CATCAATGTA TCTTATCATG TCTGGATCCC CAGGAAGCTC 9960

CTCTGTGTCC TCATAAACCC TAACCTCCTC TACTTGAGAG GACATTCCAA TCATAGGCTG 10020

CCCATCCACC CTCTGTGTCC TCCTGTTAAT TAGGTCACTT AACAAAAAGG AAATTGGGTA 10080

GGGGTTTTTC ACAGACCGCT TTCTAAGGGT AATTTTAAAA TATCTGGGAA GTCCCTTCCA 10140

CTGCTGTGTT CCAGAAGTGT TGGTAAACAG CCCACAAATG TCAACAGCAG AAACATACAA 10200

GCTGTCAGCT TTGCACAAGG GCCCAACACC CTGCTCAGCA AGAAGCACTG TGGTTGCTGT 10260

GTTAGTAATG TGCAAAACAG GAGGCACATT TTCCCCACCT GTGTAGGTTC CAAAATATCT 10320

AGTGTTTTCA TTTTTACTTG GATCAGGAAC CCAGCACTCC ACTGGATAAG CATTATCCTT 10380

ATCCAAAACA GCCTTGTGGT CAGTGTTCAT CTGCTGACTG TCAACTGTAG CATTTTTTGG 10440

GGTTACAGTT TGAGCAGGAT ATTTGGTCCT GTAGTTTGCT AACACACCCT GCAGCTCCAA 10500

AGGTTCCCCA CCAACAGCAA AAAAATGAAA ATTTGACCCT TGAATGGGTT TTCCAGCACC 10560

ATTTTCATGA GTTTTTTGTG TCCCTGAATG CAAGTTTAAC ATAGCAGTTA CCCCAATAAC 10620

CTCAGTTTTA ACAGTAACAG CTTCCCACAT CAAAATATTT CCACAGGTTA AGTCCTCATT 10680

TAAATTAGGC AAAGGAA 10697

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10549 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATTCGA CCTCGAGGGA TCTTTGTGAA GGAACCTTAC TTCTGTGGTG 2160

TGACATAATT GGACAAACTA CCTACAGAGA TTTAAAGCTC TAAGGTAAAT ATAAAATTTT 2220

TAAGTGTATA ATGTGTTAAA CTACTGATTC TAATTGTTTG TGTATTTTAG ATTCCAACCT 2280

ATGGAACTGA TGAATGGGAG CAGTGGTGGA ATGCCTTTAA TGAGGAAAAC CTGTTTTGCT 2340

CAGAAGAAAT GCCATCTAGT GATGATGAGG CTACTGCTGA CTCTCAACAT TCTACTCCTC 2400

CAAAAAAGAA GAGAAAGGTA GAAGACCCCA AGGACTTTCC TTCAGAATTG CTAAGTTTTT 2460

TGAGTCATGC TGTGTTTAGT AATAGAACTC TTGCTTGCTT TGCTATTTAC ACCACAAAGG 2520

AAAAAGCTGC ACTGCTATAC AAGAAAATTA TGGAAAAATA TTCTGTAACC TTTATAAGTA 2580

GGCATAACAG TTATAATCAT AACATACTGT TTTTTCTTAC TCCACACAGG CATAGAGTGT 2640

CTGCTATTAA TAACTATGCT CAAAAATTGT GTACCTTTAG CTTTTTAATT TGTAAAGGGG 2700

TTAATAAGGA ATATTTGATG TATAGTGCCT TGACTAGAGA TCATAATCAG CCATACCACA 2760

TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC TCCCCCTGAA CCTGAAACAT 2820

AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG CTTATAATGG TTACAAATAA 2880

AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT 2940

TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC GGCTGTGGAA TGTGTGTCAG 3000

TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC 3060

AATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC CAGCAGGCAG AAGTATGCAA 3120

AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC 3180

CTAACTCCGC CCAGTTCCGC CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT 3240

GCAGAGGCCG AGGCCGCCTC GGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT 3300

GGAGGCCTAG GCTTTTGCAA AAAGCTTCAC GCTGCCGCAA GCACTCAGGG CGCAAGGGCT 3360

GCTAAAGGAA GCGGAACACG TAGAAAGCCA GTCCGCAGAA ACGGTGCTGA CCCCGGATGA 3420

ATGTCAGCTA CTGGGCTATC TGGACAAGGG AAAACGCAAG CGCAAAGAGA AAGCAGGTAG 3480

CTTGCAGTGG GCTTACATGG CGATAGCTAG ACTGGGCGGT TTTATGGACA GCAAGCGAAC 3540

CGGAATTGCC AGCTGGGGCG CCCTCTGGTA AGGTTGGGAA GCCCTGCAAA GTAAACTGGA 3600

TGGCTTTCTT GCCGCCAAGG ATCTGATGGC GCAGGGGATC AAGATCTGAT CAAGAGACAG 3660

GATGAGGATC GTTTCGCATG ATTGAACAAG ATGGATTGCA CGCAGGTTCT CCGGCCGCTT 3720

GGGTGGAGAG GCTATTCGGC TATGACTGGG CACAACAGAC AATCGGCTGC TCTGATGCCG 3780

CCGTGTTCCG GCTGTCAGCG CAGGGGCGCC CGGTTCTTTT TGTCAAGACC GACCTGTCCG 3840

GTGCCCTGAA TGAACTGCAG GACGAGGCAG CGCGGCTATC GTGGCTGGCC ACGACGGGCG 3900

TTCCTTGCGC AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG 3960

GCGAAGTGCC GGGGCAGGAT CTCCTGTCAT CTCACCTTGC TCCTGCCGAG AAAGTATCCA 4020

TCATGGCTGA TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC 4080

ACCAAGCGAA ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC 4140

AGGATGATCT GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTC GCCAGGCTCA 4200

AGGCGCGCAT GCCCGACGGC GAGGATCTCG TCGTGACCCA TGGCGATGCC TGCTTGCCGA 4260

ATATCATGGT GGAAAATGGC CGCTTTTCTG GATTCATCGA CTGTGGCCGG CTGGGTGTGG 4320

CGGACCGCTA TCAGGACATA GCGTTGGCTA CCCGTGATAT TGCTGAAGAG CTTGGCGGCG 4380

AATGGGCTGA CCGCTTCCTC GTGCTTTACG GTATCGCCGC TCCCGATTCG CAGCGCATCG 4440

CCTTCTATCG CCTTCTTGAC GAGTTCTTCT GAGCGGGACT CTGGGGTTCG AAATGACCGA 4500

CCAAGCGACG CCCAACCTGC CATCACGAGA TTTCGATTCC ACCGCCGCCT TCTATGAAAG 4560

GTTGGGCTTC GGAATCGTTT TCCGGGACGC CGGCTGGATG ATCCTCCAGC GCGGGGATCT 4620

CATGCTGGAG TTCTTCGCCC ACCCCGGGCT CGATCCCCTC GCGAGTTGGT TCAGCTGCTG 4680

CCTGAGGCTG GACGACCTCG CGGAGTTCTA CCGGCAGTGC AAATCCGTCG GCATCCAGGA 4740

AACCAGCAGC GGCTATCCGC GCATCCATGC CCCCGAACTG CAGGAGTGGG GAGGCACGAT 4800

GGCCGCTTTG GTCCCGGATC TTTGTGAAGG AACCTTACTT CTGTGGTGTG ACATAATTGG 4860

ACAAACTACC TACAGAGATT TAAAGCTCTA AGGTAAATAT AAAATTTTTA AGTGTATAAT 4920

GTGTTAAACT ACTGATTCTA ATTGTTTGTG TATTTTAGAT TCCAACCTAT GGAACTGATG 4980

AATGGGAGCA GTGGTGGAAT GCCTTTAATG AGGAAAACCT GTTTTGCTCA GAAGAAATGC 5040

CATCTAGTGA TGATGAGGCT ACTGCTGACT CTCAACATTC TACTCCTCCA AAAAAGAAGA 5100

GAAAGGTAGA AGACCCCAAG GACTTTCCTT CAGAATTGCT AAGTTTTTTG AGTCATGCTG 5160

TGTTTAGTAA TAGAACTCTT GCTTGCTTTG CTATTTACAC CACAAAGGAA AAAGCTGCAC 5220

TGCTATACAA GAAAATTATG GAAAAATATT CTGTAACCTT TATAAGTAGG CATAACAGTT 5280

ATAATCATAA CATACTGTTT TTTCTTACTC CACACAGGCA TAGAGTGTCT GCTATTAATA 5340

ACTATGCTCA AAAATTGTGT ACCTTTAGCT TTTTAATTTG TAAAGGGGTT AATAAGGAAT 5400

ATTTGATGTA TAGTGCCTTG ACTAGAGATC ATAATCAGCC ATACCACATT TGTAGAGGTT 5460

TTACTTGCTT TAAAAAACCT CCCACACCTC CCCCTGAACC TGAAACATAA AATGAATGCA 5520

ATTGTTGTTG TTAACTTGTT TATTGCAGCT TATAATGGTT ACAAATAAAG CAATAGCATC 5580

ACAAATTTCA CAAATAAAGC ATTTTTTTCA CTGCATTCTA GTTGTGGTTT GTCCAAACTC 5640

ATCAATGTAT CTTATCATGT CTGGATCCCC AGGAAGCTCC TCTGTGTCCT CATAAACCCT 5700

AACCTCCTCT ACTTGAGAGG ACATTCCAAT CATAGGCTGC CCATCCACCC TCTGTGTCCT 5760

CCTGTTAATT AGGTCACTTA ACAAAAAGGA AATTGGGTAG GGGTTTTTCA CAGACCGCTT 5820

TCTAAGGGTA ATTTTAAAAT ATCTGGGAAG TCCCTTCCAC TGCTGTGTTC CAGAAGTGTT 5880

GGTAAACAGC CCACAAATGT CAACAGCAGA AACATACAAG CTGTCAGCTT TGCACAAGGG 5940

CCCAACACCC TGCTCATCAA GAAGCACTGT GGTTGCTGTG TTAGTAATGT GCAAAACAGG 6000

AGGCACATTT TCCCCACCTG TGTAGGTTCC AAAATATCTA GTGTTTTCAT TTTTACTTGG 6060

ATCAGGAACC CAGCACTCCA CTGGATAAGC ATTATCCTTA TCCAAAACAG CCTTGTGGTC 6120

AGTGTTCATC TGCTGACTGT CAACTGTAGC ATTTTTTGGG GTTACAGTTT GAGCAGGATA 6180

TTTGGTCCTG TAGTTTGCTA ACACACCCTG CAGCTCCAAA GGTTCCCCAC CAACAGCAAA 6240

AAAATGAAAA TTTGACCCTT GAATGGGTTT TCCAGCACCA TTTTCATGAG TTTTTTGTGT 6300

CCCTGAATGC AAGTTTAACA TAGCAGTTAC CCCAATAACC TCAGTTTTAA CAGTAACAGC 6360

TTCCCACATC AAAATATTTC CACAGGTTAA GTCCTCATTT AAATTAGGCA AAGGAATTAT 6420

ACACTCCGCT ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG 6480

CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG 6540

TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGGCAGC 6600

GGATCATAAT CAGCCATACC ACATTTGTAG AGGTTTTACT TGCTTTAAAA AACCTCCCAC 6660

ACCTCCCCCT GAACCTGAAA CATAAAATGA ATGCAATTGT TGTTGTTAAC TTGTTTATTG 6720

CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT 6780

TTTCACTGCA TTCTAGTTGT GGTTTGTCCA AACTCATCAA TGTATCTTAT CATGTCTGGA 6840

TCATAATCAG CCATACCACA TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC 6900

TCCCCCTGAA CCTGAAACAT AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG 6960

CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT 7020

CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC 7080

CAAGTTCATC TATTTCCTCC CACATCTGGT ATAAAAGGAG GCAGTGGCCC ACAGAGGAGC 7140

ACAGCTGTGT TTGGCTGCAG GGCCAAGAGC GCTGTCAAGA AGACCCACAC GCCCCCCTCC 7200

AGCAGCTGAA TTCCAGCTGG CATTCCGGTA CTGTTGGTAA AATGGAAGAC GCCAAAAACA 7260

TAAAGAAAGG CCCGGCGCCA TTCTATCCTC TAGAGGATGG AACCGCTGGA GAGCAACTGC 7320

ATAAGGCTAT GAAGAGATAC GCCCTGGTTC CTGGAACAAT TGCTTTTACA GATGCACATA 7380

TCGAGGTGAA CATCACGTAC GCGGAATACT TCGAAATGTC CGTTCGGTTG GCAGAAGCTA 7440

TGAAACGATA TGGGCTGAAT ACAAATCACA GAATCGTCGT ATGCAGTGAA AACTCTCTTC 7500

AATTCTTTAT GCCGGTGTTG GGCGCGTTAT TTATCGGAGT TGCAGTTGCG CCCGCGAACG 7560

ACATTTATAA TGAACGTGAA TTGCTCAACA GTATGAACAT TTCGCAGCCT ACCGTAGTGT 7620

TTGTTTCCAA AAAGGGGTTG CAAAAAATTT TGAACGTGCA AAAAAAATTA CCAATAATCC 7680

AGAAAATTAT TATCATGGAT TCTAAAACGG ATTACCAGGG ATTTCAGTCG ATGTACACGT 7740

TCGTCACATC TCATCTACCT CCCGGTTTTA ATGAATACGA TTTTGTACCA GAGTCCTTTG 7800

ATCGTGACAA AACAATTGCA CTGATAATGA ATTCCTCTGG ATCTACTGGG TTACCTAAGG 7860

GTGTGGCCCT TCCGCATAGA ACTGCCTGCG TCAGATTCTC GCATGCCAGA GATCCTATTT 7920

TTGGCAATCA AATCATTCCG GATACTGCGA TTTTAAGTGT TGTTCCATTC CATCACGGTT 7980

TTGGAATGTT TACTACACTC GGATATTTGA TATGTGGATT TCGAGTCGTC TTAATGTATA 8040

GATTTGAAGA AGAGCTGTTT TTACGATCCC TTCAGGATTA CAAAATTCAA AGTGCGTTGC 8100

TAGTACCAAC CCTATTTTCA TTCTTCGCCA AAAGCACTCT GATTGACAAA TACGATTTAT 8160

CTAATTTACA CGAAATTGCT TCTGGGGGCG CACCTCTTTC GAAAGAAGTC GGGGAAGCGG 8220

TTGCAAAACG CTTCCATCTT CCAGGGATAC GACAAGGATA TGGGCTCACT GAGACTACAT 8280

CAGCTATTCT GATTACACCC GAGGGGGATG ATAAACCGGG CGCGGTCGGT AAAGTTGTTC 8340

CATTTTTTGA AGCGAAGGTT GTGGATCTGG ATACCGGGAA AACGCTGGGC GTTAATCAGA 8400

GAGGCGAATT ATGTGTCAGA GGACCTATGA TTATGTCCGG TTATGTAAAC AATCCGGAAG 8460

CGACCAACGC CTTGATTGAC AAGGATGGAT GGCTACATTC TGGAGACATA GCTTACTGGG 8520

ACGAAGACGA ACACTTCTTC ATAGTTGACC GCTTGAAGTC TTTAATTAAA TACAAAGGAT 8580

ATCAGGTGGC CCCCGCTGAA TTGGAATCGA TATTGTTACA ACACCCCAAC ATCTTCGACG 8640

CGGGCGTGGC AGGTCTTCCC GACGATGACG CCGGTGAACT TCCCGCCGCC GTTGTTGTTT 8700

TGGAGCACGG AAAGACGATG ACGGAAAAAG AGATCGTGGA TTACGTCGCC AGTCAAGTAA 8760

CAACCGCGAA AAAGTTGCGC GGAGGAGTTG TGTTTGTGGA CGAAGTACCG AAAGGTCTTA 8820

CCGGAAAACT CGACGCAAGA AAAATCAGAG AGATCCTCAT AAAGGCCAAG AAGGGCGGAA 8880

AGTCCAAATT GTAAAATGTA ACTGTATTCA GCGATGACGA AATTCTTAGC TATTGTAATG 8940

ACTCTAGAGG ATCTTTGTGA AGGAACCTTA CTTCTGTGGT GTGACATAAT TGGACAAACT 9000

ACCTACAGAG ATTTAAAGCT CTAAGGTAAA TATAAAATTT TTAAGTGTAT AATGTGTTAA 9060

ACTACTGATT CTAATTGTTT GTGTATTTTA GATTCCAACC TATGGAACTG ATGAATGGGA 9120

GCAGTGGTGG AATGCCTTTA ATGAGGAAAA CCTGTTTTGC TCAGAAGAAA TGCCATCTAG 9180

TGATGATGAG GCTACTGCTG ACTCTCAACA TTCTACTCCT CCAAAAAAGA AGAGAAAGGT 9240

AGAAGACCCC AAGGACTTTC CTTCAGAATT GCTAAGTTTT TTGAGTCATG CTGTGTTTAG 9300

TAATAGAACT CTTGCTTGCT TTGCTATTTA CACCACAAAG GAAAAAGCTG CACTGCTATA 9360

CAAGAAAATT ATGGAAAAAT ATTCTGTAAC CTTTATAAGT AGGCATAACA GTTATAATCA 9420

TAACATACTG TTTTTTCTTA CTCCACACAG GCATAGAGTG TCTGCTATTA ATAACTATGC 9480

TCAAAAATTG TGTACCTTTA GCTTTTTAAT TTGTAAAGGG GTTAATAAGG AATATTTGAT 9540

GTATAGTGCC TTGACTAGAG ATCATAATCA GCCATACCAC ATTTGTAGAG GTTTTACTTG 9600

CTTTAAAAAA CCTCCCACAC CTCCCCCTGA ACCTGAAACA TAAAATGAAT GCAATTGTTG 9660

TTGTTAACTT GTTTATTGCA GCTTATAATG GTTACAAATA AAGCAATAGC ATCACAAATT 9720

TCACAAATAA AGCATTTTTT TCACTGCATT CTAGTTGTGG TTTGTCCAAA CTCATCAATG 9780

TATCTTATCA TGTCTGGATC CCCAGGAAGC TCCTCTGTGT CCTCATAAAC CCTAACCTCC 9840

TCTACTTGAG AGGACATTCC AATCATAGGC TGCCCATCCA CCCTCTGTGT CCTCCTGTTA 9900

ATTAGGTCAC TTAACAAAAA GGAAATTGGG TAGGGGTTTT TCACAGACCG CTTTCTAAGG 9960

GTAATTTTAA AATATCTGGG AAGTCCCTTC CACTGCTGTG TTCCAGAAGT GTTGGTAAAC 10020

AGCCCACAAA TGTCAACAGC AGAAACATAC AAGCTGTCAG CTTTGCACAA GGGCCCAACA 10080

CCCTGCTCAG CAAGAAGCAC TGTGGTTGCT GTGTTAGTAA TGTGCAAAAC AGGAGGCACA 10140

TTTTCCCCAC CTGTGTAGGT TCCAAAATAT CTAGTGTTTT CATTTTTACT TGGATCAGGA 10200

ACCCAGCACT CCACTGGATA AGCATTATCC TTATCCAAAA CAGCCTTGTG GTCAGTGTTC 10260

ATCTGCTGAC TGTCAACTGT AGCATTTTTT GGGGTTACAG TTTGAGCAGG ATATTTGGTC 10320

CTGTAGTTTG CTAACACACC CTGCAGCTCC AAAGGTTCCC CACCAACAGC AAAAAAATGA 10380

AAATTTGACC CTTGAATGGG TTTTCCAGCA CCATTTTCAT GAGTTTTTTG TGTCCCTGAA 10440

TGCAAGTTTA ACATAGCAGT TACCCCAATA ACCTCAGTTT TAACAGTAAC AGCTTCCCAC 10500

ATCAAAATAT TTCCACAGGT TAAGTCCTCA TTTAAATTAG GCAAAGGAA 10549

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10558 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATTCGA CCTCGAGGGA TCTTTGTGAA GGAACCTTAC TTCTGTGGTG 2160

TGACATAATT GGACAAACTA CCTACAGAGA TTTAAAGCTC TAAGGTAAAT ATAAAATTTT 2220

TAAGTGTATA ATGTGTTAAA CTACTGATTC TAATTGTTTG TGTATTTTAG ATTCCAACCT 2280

ATGGAACTGA TGAATGGGAG CAGTGGTGGA ATGCCTTTAA TGAGGAAAAC CTGTTTTGCT 2340

CAGAAGAAAT GCCATCTAGT GATGATGAGG CTACTGCTGA CTCTCAACAT TCTACTCCTC 2400

CAAAAAAGAA GAGAAAGGTA GAAGACCCCA AGGACTTTCC TTCAGAATTG CTAAGTTTTT 2460

TGAGTCATGC TGTGTTTAGT AATAGAACTC TTGCTTGCTT TGCTATTTAC ACCACAAAGG 2520

AAAAAGCTGC ACTGCTATAC AAGAAAATTA TGGAAAAATA TTCTGTAACC TTTATAAGTA 2580

GGCATAACAG TTATAATCAT AACATACTGT TTTTTCTTAC TCCACACAGG CATAGAGTGT 2640

CTGCTATTAA TAACTATGCT CAAAAATTGT GTACCTTTAG CTTTTTAATT TGTAAAGGGG 2700

TTAATAAGGA ATATTTGATG TATAGTGCCT TGACTAGAGA TCATAATCAG CCATACCACA 2760

TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC TCCCCCTGAA CCTGAAACAT 2820

AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG CTTATAATGG TTACAAATAA 2880

AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT 2940

TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC GGCTGTGGAA TGTGTGTCAG 3000

TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC 3060

AATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC CAGCAGGCAG AAGTATGCAA 3120

AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC 3180

CTAACTCCGC CCAGTTCCGC CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT 3240

GCAGAGGCCG AGGCCGCCTC GGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT 3300

GGAGGCCTAG GCTTTTGCAA AAAGCTTCAC GCTGCCGCAA GCACTCAGGG CGCAAGGGCT 3360

GCTAAAGGAA GCGGAACACG TAGAAAGCCA GTCCGCAGAA ACGGTGCTGA CCCCGGATGA 3420

ATGTCAGCTA CTGGGCTATC TGGACAAGGG AAAACGCAAG CGCAAAGAGA AAGCAGGTAG 3480

CTTGCAGTGG GCTTACATGG CGATAGCTAG ACTGGGCGGT TTTATGGACA GCAAGCGAAC 3540

CGGAATTGCC AGCTGGGGCG CCCTCTGGTA AGGTTGGGAA GCCCTGCAAA GTAAACTGGA 3600

TGGCTTTCTT GCCGCCAAGG ATCTGATGGC GCAGGGGATC AAGATCTGAT CAAGAGACAG 3660

GATGAGGATC GTTTCGCATG ATTGAACAAG ATGGATTGCA CGCAGGTTCT CCGGCCGCTT 3720

GGGTGGAGAG GCTATTCGGC TATGACTGGG CACAACAGAC AATCGGCTGC TCTGATGCCG 3780

CCGTGTTCCG GCTGTCAGCG CAGGGGCGCC CGGTTCTTTT TGTCAAGACC GACCTGTCCG 3840

GTGCCCTGAA TGAACTGCAG GACGAGGCAG CGCGGCTATC GTGGCTGGCC ACGACGGGCG 3900

TTCCTTGCGC AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG 3960

GCGAAGTGCC GGGGCAGGAT CTCCTGTCAT CTCACCTTGC TCCTGCCGAG AAAGTATCCA 4020

TCATGGCTGA TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC 4080

ACCAAGCGAA ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC 4140

AGGATGATCT GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTC GCCAGGCTCA 4200

AGGCGCGCAT GCCCGACGGC GAGGATCTCG TCGTGACCCA TGGCGATGCC TGCTTGCCGA 4260

ATATCATGGT GGAAAATGGC CGCTTTTCTG GATTCATCGA CTGTGGCCGG CTGGGTGTGG 4320

CGGACCGCTA TCAGGACATA GCGTTGGCTA CCCGTGATAT TGCTGAAGAG CTTGGCGGCG 4380

AATGGGCTGA CCGCTTCCTC GTGCTTTACG GTATCGCCGC TCCCGATTCG CAGCGCATCG 4440

CCTTCTATCG CCTTCTTGAC GAGTTCTTCT GAGCGGGACT CTGGGGTTCG AAATGACCGA 4500

CCAAGCGACG CCCAACCTGC CATCACGAGA TTTCGATTCC ACCGCCGCCT TCTATGAAAG 4560

GTTGGGCTTC GGAATCGTTT TCCGGGACGC CGGCTGGATG ATCCTCCAGC GCGGGGATCT 4620

CATGCTGGAG TTCTTCGCCC ACCCCGGGCT CGATCCCCTC GCGAGTTGGT TCAGCTGCTG 4680

CCTGAGGCTG GACGACCTCG CGGAGTTCTA CCGGCAGTGC AAATCCGTCG GCATCCAGGA 4740

AACCAGCAGC GGCTATCCGC GCATCCATGC CCCCGAACTG CAGGAGTGGG GAGGCACGAT 4800

GGCCGCTTTG GTCCCGGATC TTTGTGAAGG AACCTTACTT CTGTGGTGTG ACATAATTGG 4860

ACAAACTACC TACAGAGATT TAAAGCTCTA AGGTAAATAT AAAATTTTTA AGTGTATAAT 4920

GTGTTAAACT ACTGATTCTA ATTGTTTGTG TATTTTAGAT TCCAACCTAT GGAACTGATG 4980

AATGGGAGCA GTGGTGGAAT GCCTTTAATG AGGAAAACCT GTTTTGCTCA GAAGAAATGC 5040

CATCTAGTGA TGATGAGGCT ACTGCTGACT CTCAACATTC TACTCCTCCA AAAAAGAAGA 5100

GAAAGGTAGA AGACCCCAAG GACTTTCCTT CAGAATTGCT AAGTTTTTTG AGTCATGCTG 5160

TGTTTAGTAA TAGAACTCTT GCTTGCTTTG CTATTTACAC CACAAAGGAA AAAGCTGCAC 5220

TGCTATACAA GAAAATTATG GAAAAATATT CTGTAACCTT TATAAGTAGG CATAACAGTT 5280

ATAATCATAA CATACTGTTT TTTCTTACTC CACACAGGCA TAGAGTGTCT GCTATTAATA 5340

ACTATGCTCA AAAATTGTGT ACCTTTAGCT TTTTAATTTG TAAAGGGGTT AATAAGGAAT 5400

ATTTGATGTA TAGTGCCTTG ACTAGAGATC ATAATCAGCC ATACCACATT TGTAGAGGTT 5460

TTACTTGCTT TAAAAAACCT CCCACACCTC CCCCTGAACC TGAAACATAA AATGAATGCA 5520

ATTGTTGTTG TTAACTTGTT TATTGCAGCT TATAATGGTT ACAAATAAAG CAATAGCATC 5580

ACAAATTTCA CAAATAAAGC ATTTTTTTCA CTGCATTCTA GTTGTGGTTT GTCCAAACTC 5640

ATCAATGTAT CTTATCATGT CTGGATCCCC AGGAAGCTCC TCTGTGTCCT CATAAACCCT 5700

AACCTCCTCT ACTTGAGAGG ACATTCCAAT CATAGGCTGC CCATCCACCC TCTGTGTCCT 5760

CCTGTTAATT AGGTCACTTA ACAAAAAGGA AATTGGGTAG GGGTTTTTCA CAGACCGCTT 5820

TCTAAGGGTA ATTTTAAAAT ATCTGGGAAG TCCCTTCCAC TGCTGTGTTC CAGAAGTGTT 5880

GGTAAACAGC CCACAAATGT CAACAGCAGA AACATACAAG CTGTCAGCTT TGCACAAGGG 5940

CCCAACACCC TGCTCATCAA GAAGCACTGT GGTTGCTGTG TTAGTAATGT GCAAAACAGG 6000

AGGCACATTT TCCCCACCTG TGTAGGTTCC AAAATATCTA GTGTTTTCAT TTTTACTTGG 6060

ATCAGGAACC CAGCACTCCA CTGGATAAGC ATTATCCTTA TCCAAAACAG CCTTGTGGTC 6120

AGTGTTCATC TGCTGACTGT CAACTGTAGC ATTTTTTGGG GTTACAGTTT GAGCAGGATA 6180

TTTGGTCCTG TAGTTTGCTA ACACACCCTG CAGCTCCAAA GGTTCCCCAC CAACAGCAAA 6240

AAAATGAAAA TTTGACCCTT GAATGGGTTT TCCAGCACCA TTTTCATGAG TTTTTTGTGT 6300

CCCTGAATGC AAGTTTAACA TAGCAGTTAC CCCAATAACC TCAGTTTTAA CAGTAACAGC 6360

TTCCCACATC AAAATATTTC CACAGGTTAA GTCCTCATTT AAATTAGGCA AAGGAATTAT 6420

ACACTCCGCT ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG 6480

CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG 6540

TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGGCAGC 6600

GGATCATAAT CAGCCATACC ACATTTGTAG AGGTTTTACT TGCTTTAAAA AACCTCCCAC 6660

ACCTCCCCCT GAACCTGAAA CATAAAATGA ATGCAATTGT TGTTGTTAAC TTGTTTATTG 6720

CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT 6780

TTTCACTGCA TTCTAGTTGT GGTTTGTCCA AACTCATCAA TGTATCTTAT CATGTCTGGA 6840

TCATAATCAG CCATACCACA TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC 6900

TCCCCCTGAA CCTGAAACAT AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG 6960

CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT 7020

CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC 7080

CAGTGGGGAG TCAGCCGTGT ATCATCGCCC ACATCTGGTA TAAAAGGAGG CAGTGGCCCA 7140

CAGAGGAGCA CAGCTGTGTT TGGCTGCAGG GCCAAGAGCG CTGTCAAGAA GACCCACACG 7200

CCCCCCTCCA GCAGCTGAAT TCCAGCTGGC ATTCCGGTAC TGTTGGTAAA ATGGAAGACG 7260

CCAAAAACAT AAAGAAAGGC CCGGCGCCAT TCTATCCTCT AGAGGATGGA ACCGCTGGAG 7320

AGCAACTGCA TAAGGCTATG AAGAGATACG CCCTGGTTCC TGGAACAATT GCTTTTACAG 7380

ATGCACATAT CGAGGTGAAC ATCACGTACG CGGAATACTT CGAAATGTCC GTTCGGTTGG 7440

CAGAAGCTAT GAAACGATAT GGGCTGAATA CAAATCACAG AATCGTCGTA TGCAGTGAAA 7500

ACTCTCTTCA ATTCTTTATG CCGGTGTTGG GCGCGTTATT TATCGGAGTT GCAGTTGCGC 7560

CCGCGAACGA CATTTATAAT GAACGTGAAT TGCTCAACAG TATGAACATT TCGCAGCCTA 7620

CCGTAGTGTT TGTTTCCAAA AAGGGGTTGC AAAAAATTTT GAACGTGCAA AAAAAATTAC 7680

CAATAATCCA GAAAATTATT ATCATGGATT CTAAAACGGA TTACCAGGGA TTTCAGTCGA 7740

TGTACACGTT CGTCACATCT CATCTACCTC CGGGTTTTAA TGAATACGAT TTTGTACCAG 7800

AGTCCTTTGA TCGTGACAAA ACAATTGCAC TGATAATGAA TTCCTCTGGA TCTACTGGGT 7860

TACCTAAGGG TGTGGCCCTT CCGCATAGAA CTGCCTGCGT CAGATTCTCG CATGCCAGAG 7920

ATCCTATTTT TGGCAATCAA ATCATTCCGG ATACTGCGAT TTTAAGTGTT GTTCCATTCC 7980

ATCACGGTTT TGGAATGTTT ACTACACTCG GATATTTGAT ATGTGGATTT CGAGTCGTCT 8040

TAATGTATAG ATTTGAAGAA GAGCTGTTTT TACGATCCCT TCAGGATTAC AAAATTCAAA 8100

GTGCGTTGCT AGTACCAACC CTATTTTCAT TCTTCGCCAA AAGCACTCTG ATTGACAAAT 8160

ACGATTTATC TAATTTACAC GAAATTGCTT CTGGGGGCGC ACCTCTTTCG AAAGAAGTCG 8220

GGGAAGCGGT TGCAAAACGC TTCCATCTTC CAGGGATACG ACAAGGATAT GGGCTCACTG 8280

AGACTACATC AGCTATTCTG ATTACACCCG AGGGGGATGA TAAACCGGGC GCGGTCGGTA 8340

AAGTTGTTCC ATTTTTTGAA GCGAAGGTTG TGGATCTGGA TACCGGGAAA ACGCTGGGCG 8400

TTAATCAGAG AGGCGAATTA TGTGTCAGAG GACCTATGAT TATGTCCGGT TATGTAAACA 8460

ATCCGGAAGC GACCAACGCC TTGATTGACA AGGATGGATG GCTACATTCT GGAGACATAG 8520

CTTACTGGGA CGAAGACGAA CACTTCTTCA TAGTTGACCG CTTGAAGTCT TTAATTAAAT 8580

ACAAAGGATA TCAGGTGGCC CCCGCTGAAT TGGAATCGAT ATTGTTACAA CACCCCAACA 8640

TCTTCGACGC GGGCGTGGCA GGTCTTCCCG ACGATGACGC CGGTGAACTT CCCGCCGCCG 8700

TTGTTGTTTT GGAGCACGGA AAGACGATGA CGGAAAAAGA GATCGTGGAT TACGTCGCCA 8760

GTCAAGTAAC AACCGCGAAA AAGTTGCGCG GAGGAGTTGT GTTTGTGGAC GAAGTACCGA 8820

AAGGTCTTAC CGGAAAACTC GACGCAAGAA AAATCAGAGA GATCCTCATA AAGGCCAAGA 8880

AGGGCGGAAA GTCCAAATTG TAAAATGTAA CTGTATTCAG CGATGACGAA ATTCTTAGCT 8940

ATTGTAATGA CTCTAGAGGA TCTTTGTGAA GGAACCTTAC TTCTGTGGTG TGACATAATT 9000

GGACAAACTA CCTACAGAGA TTTAAAGCTC TAAGGTAAAT ATAAAATTTT TAAGTGTATA 9060

ATGTGTTAAA CTACTGATTC TAATTGTTTG TGTATTTTAG ATTCCAACCT ATGGAACTGA 9120

TGAATGGGAG CAGTGGTGGA ATGCCTTTAA TGAGGAAAAC CTGTTTTGCT CAGAAGAAAT 9180

GCCATCTAGT GATGATGAGG CTACTGCTGA CTCTCAACAT TCTACTCCTC CAAAAAAGAA 9240

GAGAAAGGTA GAAGACCCCA AGGACTTTCC TTCAGAATTG CTAAGTTTTT TGAGTCATGC 9300

TGTGTTTAGT AATAGAACTC TTGCTTGCTT TGCTATTTAC ACCACAAAGG AAAAAGCTGC 9360

ACTGCTATAC AAGAAAATTA TGGAAAAATA TTCTGTAACC TTTATAAGTA GGCATAACAG 9420

TTATAATCAT AACATACTGT TTTTTCTTAC TCCACACAGG CATAGAGTGT CTGCTATTAA 9480

TAACTATGCT CAAAAATTGT GTACCTTTAG CTTTTTAATT TGTAAAGGGG TTAATAAGGA 9540

ATATTTGATG TATAGTGCCT TGACTAGAGA TCATAATCAG CCATACCACA TTTGTAGAGG 9600

TTTTACTTGC TTTAAAAAAC CTCCCACACC TCCCCCTGAA CCTGAAACAT AAAATGAATG 9660

CAATTGTTGT TGTTAACTTG TTTATTGCAG CTTATAATGG TTACAAATAA AGCAATAGCA 9720

TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT TTGTCCAAAC 9780

TCATCAATGT ATCTTATCAT GTCTGGATCC CCAGGAAGCT CCTCTGTGTC CTCATAAACC 9840

CTAACCTCCT CTACTTGAGA GGACATTCCA ATCATAGGCT GCCCATCCAC CCTCTGTGTC 9900

CTCCTGTTAA TTAGGTCACT TAACAAAAAG GAAATTGGGT AGGGGTTTTT CACAGACCGC 9960

TTTCTAAGGG TAATTTTAAA ATATCTGGGA AGTCCCTTCC ACTGCTGTGT TCCAGAAGTG 10020

TTGGTAAACA GCCCACAAAT GTCAACAGCA GAAACATACA AGCTGTCAGC TTTGCACAAG 10080

GGCCCAACAC CCTGCTCAGC AAGAAGCACT GTGGTTGCTG TGTTAGTAAT GTGCAAAACA 10140

GGAGGCACAT TTTCCCCACC TGTGTAGGTT CCAAAATATC TAGTGTTTTC ATTTTTACTT 10200

GGATCAGGAA CCCAGCACTC CACTGGATAA GCATTATCCT TATCCAAAAC AGCCTTGTGG 10260

TCAGTGTTCA TCTGCTGACT GTCAACTGTA GCATTTTTTG GGGTTACAGT TTGAGCAGGA 10320

TATTTGGTCC TGTAGTTTGC TAACACACCC TGCAGCTCCA AAGGTTCCCC ACCAACAGCA 10380

AAAAAATGAA AATTTGACCC TTGAATGGGT TTTCCAGCAC CATTTTCATG AGTTTTTTGT 10440

GTCCCTGAAT GCAAGTTTAA CATAGCAGTT ACCCCAATAA CCTCAGTTTT AACAGTAACA 10500

GCTTCCCACA TCAAAATATT TCCACAGGTT AAGTCCTCAT TTAAATTAGG CAAAGGAA 10558

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10569 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATTCGA CCTCGAGGGA TCTTTGTGAA GGAACCTTAC TTCTGTGGTG 2160

TGACATAATT GGACAAACTA CCTACAGAGA TTTAAAGCTC TAAGGTAAAT ATAAAATTTT 2220

TAAGTGTATA ATGTGTTAAA CTACTGATTC TAATTGTTTG TGTATTTTAG ATTCCAACCT 2280

ATGGAACTGA TGAATGGGAG CAGTGGTGGA ATGCCTTTAA TGAGGAAAAC CTGTTTTGCT 2340

CAGAAGAAAT GCCATCTAGT GATGATGAGG CTACTGCTGA CTCTCAACAT TCTACTCCTC 2400

CAAAAAAGAA GAGAAAGGTA GAAGACCCCA AGGACTTTCC TTCAGAATTG CTAAGTTTTT 2460

TGAGTCATGC TGTGTTTAGT AATAGAACTC TTGCTTGCTT TGCTATTTAC ACCACAAAGG 2520

AAAAAGCTGC ACTGCTATAC AAGAAAATTA TGGAAAAATA TTCTGTAACC TTTATAAGTA 2580

GGCATAACAG TTATAATCAT AACATACTGT TTTTTCTTAC TCCACACAGG CATAGAGTGT 2640

CTGCTATTAA TAACTATGCT CAAAAATTGT GTACCTTTAG CTTTTTAATT TGTAAAGGGG 2700

TTAATAAGGA ATATTTGATG TATAGTGCCT TGACTAGAGA TCATAATCAG CCATACCACA 2760

TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC TCCCCCTGAA CCTGAAACAT 2820

AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG CTTATAATGG TTACAAATAA 2880

AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT 2940

TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC GGCTGTGGAA TGTGTGTCAG 3000

TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC 3060

AATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC CAGCAGGCAG AAGTATGCAA 3120

AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC 3180

CTAACTCCGC CCAGTTCCGC CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT 3240

GCAGAGGCCG AGGCCGCCTC GGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT 3300

GGAGGCCTAG GCTTTTGCAA AAAGCTTCAC GCTGCCGCAA GCACTCAGGG CGCAAGGGCT 3360

GCTAAAGGAA GCGGAACACG TAGAAAGCCA GTCCGCAGAA ACGGTGCTGA CCCCGGATGA 3420

ATGTCAGCTA CTGGGCTATC TGGACAAGGG AAAACGCAAG CGCAAAGAGA AAGCAGGTAG 3480

CTTGCAGTGG GCTTACATGG CGATAGCTAG ACTGGGCGGT TTTATGGACA GCAAGCGAAC 3540

CGGAATTGCC AGCTGGGGCG CCCTCTGGTA AGGTTGGGAA GCCCTGCAAA GTAAACTGGA 3600

TGGCTTTCTT GCCGCCAAGG ATCTGATGGC GCAGGGGATC AAGATCTGAT CAAGAGACAG 3660

GATGAGGATC GTTTCGCATG ATTGAACAAG ATGGATTGCA CGCAGGTTCT CCGGCCGCTT 3720

GGGTGGAGAG GCTATTCGGC TATGACTGGG CACAACAGAC AATCGGCTGC TCTGATGCCG 3780

CCGTGTTCCG GCTGTCAGCG CAGGGGCGCC CGGTTCTTTT TGTCAAGACC GACCTGTCCG 3840

GTGCCCTGAA TGAACTGCAG GACGAGGCAG CGCGGCTATC GTGGCTGGCC ACGACGGGCG 3900

TTCCTTGCGC AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG 3960

GCGAAGTGCC GGGGCAGGAT CTCCTGTCAT CTCACCTTGC TCCTGCCGAG AAAGTATCCA 4020

TCATGGCTGA TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC 4080

ACCAAGCGAA ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC 4140

AGGATGATCT GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTC GCCAGGCTCA 4200

AGGCGCGCAT GCCCGACGGC GAGGATCTCG TCGTGACCCA TGGCGATGCC TGCTTGCCGA 4260

ATATCATGGT GGAAAATGGC CGCTTTTCTG GATTCATCGA CTGTGGCCGG CTGGGTGTGG 4320

CGGACCGCTA TCAGGACATA GCGTTGGCTA CCCGTGATAT TGCTGAAGAG CTTGGCGGCG 4380

AATGGGCTGA CCGCTTCCTC GTGCTTTACG GTATCGCCGC TCCCGATTCG CAGCGCATCG 4440

CCTTCTATCG CCTTCTTGAC GAGTTCTTCT GAGCGGGACT CTGGGGTTCG AAATGACCGA 4500

CCAAGCGACG CCCAACCTGC CATCACGAGA TTTCGATTCC ACCGCCGCCT TCTATGAAAG 4560

GTTGGGCTTC GGAATCGTTT TCCGGGACGC CGGCTGGATG ATCCTCCAGC GCGGGGATCT 4620

CATGCTGGAG TTCTTCGCCC ACCCCGGGCT CGATCCCCTC GCGAGTTGGT TCAGCTGCTG 4680

CCTGAGGCTG GACGACCTCG CGGAGTTCTA CCGGCAGTGC AAATCCGTCG GCATCCAGGA 4740

AACCAGCAGC GGCTATCCGC GCATCCATGC CCCCGAACTG CAGGAGTGGG GAGGCACGAT 4800

GGCCGCTTTG GTCCCGGATC TTTGTGAAGG AACCTTACTT CTGTGGTGTG ACATAATTGG 4860

ACAAACTACC TACAGAGATT TAAAGCTCTA AGGTAAATAT AAAATTTTTA AGTGTATAAT 4920

GTGTTAAACT ACTGATTCTA ATTGTTTGTG TATTTTAGAT TCCAACCTAT GGAACTGATG 4980

AATGGGAGCA GTGGTGGAAT GCCTTTAATG AGGAAAACCT GTTTTGCTCA GAAGAAATGC 5040

CATCTAGTGA TGATGAGGCT ACTGCTGACT CTCAACATTC TACTCCTCCA AAAAAGAAGA 5100

GAAAGGTAGA AGACCCCAAG GACTTTCCTT CAGAATTGCT AAGTTTTTTG AGTCATGCTG 5160

TGTTTAGTAA TAGAACTCTT GCTTGCTTTG CTATTTACAC CACAAAGGAA AAAGCTGCAC 5220

TGCTATACAA GAAAATTATG GAAAAATATT CTGTAACCTT TATAAGTAGG CATAACAGTT 5280

ATAATCATAA CATACTGTTT TTTCTTACTC CACACAGGCA TAGAGTGTCT GCTATTAATA 5340

ACTATGCTCA AAAATTGTGT ACCTTTAGCT TTTTAATTTG TAAAGGGGTT AATAAGGAAT 5400

ATTTGATGTA TAGTGCCTTG ACTAGAGATC ATAATCAGCC ATACCACATT TGTAGAGGTT 5460

TTACTTGCTT TAAAAAACCT CCCACACCTC CCCCTGAACC TGAAACATAA AATGAATGCA 5520

ATTGTTGTTG TTAACTTGTT TATTGCAGCT TATAATGGTT ACAAATAAAG CAATAGCATC 5580

ACAAATTTCA CAAATAAAGC ATTTTTTTCA CTGCATTCTA GTTGTGGTTT GTCCAAACTC 5640

ATCAATGTAT CTTATCATGT CTGGATCCCC AGGAAGCTCC TCTGTGTCCT CATAAACCCT 5700

AACCTCCTCT ACTTGAGAGG ACATTCCAAT CATAGGCTGC CCATCCACCC TCTGTGTCCT 5760

CCTGTTAATT AGGTCACTTA ACAAAAAGGA AATTGGGTAG GGGTTTTTCA CAGACCGCTT 5820

TCTAAGGGTA ATTTTAAAAT ATCTGGGAAG TCCCTTCCAC TGCTGTGTTC CAGAAGTGTT 5880

GGTAAACAGC CCACAAATGT CAACAGCAGA AACATACAAG CTGTCAGCTT TGCACAAGGG 5940

CCCAACACCC TGCTCATCAA GAAGCACTGT GGTTGCTGTG TTAGTAATGT GCAAAACAGG 6000

AGGCACATTT TCCCCACCTG TGTAGGTTCC AAAATATCTA GTGTTTTCAT TTTTACTTGG 6060

ATCAGGAACC CAGCACTCCA CTGGATAAGC ATTATCCTTA TCCAAAACAG CCTTGTGGTC 6120

AGTGTTCATC TGCTGACTGT CAACTGTAGC ATTTTTTGGG GTTACAGTTT GAGCAGGATA 6180

TTTGGTCCTG TAGTTTGCTA ACACACCCTG CAGCTCCAAA GGTTCCCCAC CAACAGCAAA 6240

AAAATGAAAA TTTGACCCTT GAATGGGTTT TCCAGCACCA TTTTCATGAG TTTTTTGTGT 6300

CCCTGAATGC AAGTTTAACA TAGCAGTTAC CCCAATAACC TCAGTTTTAA CAGTAACAGC 6360

TTCCCACATC AAAATATTTC CACAGGTTAA GTCCTCATTT AAATTAGGCA AAGGAATTAT 6420

ACACTCCGCT ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG 6480

CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG 6540

TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGGCAGC 6600

GGATCATAAT CAGCCATACC ACATTTGTAG AGGTTTTACT TGCTTTAAAA AACCTCCCAC 6660

ACCTCCCCCT GAACCTGAAA CATAAAATGA ATGCAATTGT TGTTGTTAAC TTGTTTATTG 6720

CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT 6780

TTTCACTGCA TTCTAGTTGT GGTTTGTCCA AACTCATCAA TGTATCTTAT CATGTCTGGA 6840

TCATAATCAG CCATACCACA TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC 6900

TCCCCCTGAA CCTGAAACAT AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG 6960

CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT 7020

CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC 7080

CACTCCAACC TCAGCCAGAC AAGGTTGTTG ACACAAGACC CACATCTGGT ATAAAAGGAG 7140

GCAGTGGCCC ACAGAGGAGC ACAGCTGTGT TTGGCTGCAG GGCCAAGAGC GCTGTCAAGA 7200

AGACCCACAC GCCCCCCTCC AGCAGCTGAA TTCCAGCTGG CATTCCGGTA CTGTTGGTAA 7260

AATGGAAGAC GCCAAAAACA TAAAGAAAGG CCCGGCGCCA TTCTATCCTC TAGAGGATGG 7320

AACCGCTGGA GAGCAACTGC ATAAGGCTAT GAAGAGATAC GCCCTGGTTC CTGGAACAAT 7380

TGCTTTTACA GATGCACATA TCGAGGTGAA CATCACGTAC GCGGAATACT TCGAAATGTC 7440

CGTTCGGTTG GCAGAAGCTA TGAAACGATA TGGGCTGAAT ACAAATCACA GAATCGTCGT 7500

ATGCAGTGAA AACTCTCTTC AATTCTTTAT GCCGGTGTTG GGCGCGTTAT TTATCGGAGT 7560

TGCAGTTGCG CCCGCGAACG ACATTTATAA TGAACGTGAA TTGCTCAACA GTATGAACAT 7620

TTCGCAGCCT ACCGTAGTGT TTGTTTCCAA AAAGGGGTTG CAAAAAATTT TGAACGTGCA 7680

AAAAAAATTA CCAATAATCC AGAAAATTAT TATCATGGAT TCTAAAACGG ATTACCAGGG 7740

ATTTCAGTCG ATGTACACGT TCGTCACATC TCATCTACCT CCCGGTTTTA ATGAATACGA 7800

TTTTGTACCA GAGTCCTTTG ATCGTGACAA AACAATTGCA CTGATAATGA ATTCCTCTGG 7860

ATCTACTGGG TTACCTAAGG GTGTGGCCCT TCCGCATAGA ACTGCCTGCG TCAGATTCTC 7920

GCATGCCAGA GATCCTATTT TTGGCAATCA AATCATTCCG GATACTGCGA TTTTAAGTGT 7980

TGTTCCATTC CATCACGGTT TTGGAATGTT TACTACACTC GGATATTTGA TATGTGGATT 8040

TCGAGTCGTC TTAATGTATA GATTTGAAGA AGAGCTGTTT TTACGATCCC TTCAGGATTA 8100

CAAAATTCAA AGTGCGTTGC TAGTACCAAC CCTATTTTCA TTCTTCGCCA AAAGCACTCT 8160

GATTGACAAA TACGATTTAT CTAATTTACA CGAAATTGCT TCTGGGGGCG CACCTCTTTC 8220

GAAAGAAGTC GGGGAAGCGG TTGCAAAACG CTTCCATCTT CCAGGGATAC GACAAGGATA 8280

TGGGCTCACT GAGACTACAT CAGCTATTCT GATTACACCC GAGGGGGATG ATAAACCGGG 8340

CGCGGTCGGT AAAGTTGTTC CATTTTTTGA AGCGAAGGTT GTGGATCTGG ATACCGGGAA 8400

AACGCTGGGC GTTAATCAGA GAGGCGAATT ATGTGTCAGA GGACCTATGA TTATGTCCGG 8460

TTATGTAAAC AATCCGGAAG CGACCAACGC CTTGATTGAC AAGGATGGAT GGCTACATTC 8520

TGGAGACATA GCTTACTGGG ACGAAGACGA ACACTTCTTC ATAGTTGACC GCTTGAAGTC 8580

TTTAATTAAA TACAAAGGAT ATCAGGTGGC CCCCGCTGAA TTGGAATCGA TATTGTTACA 8640

ACACCCCAAC ATCTTCGACG CGGGCGTGGC AGGTCTTCCC GACGATGACG CCGGTGAACT 8700

TCCCGCCGCC GTTGTTGTTT TGGAGCACGG AAAGACGATG ACGGAAAAAG AGATCGTGGA 8760

TTACGTCGCC AGTCAAGTAA CAACCGCGAA AAAGTTGCGC GGAGGAGTTG TGTTTGTGGA 8820

CGAAGTACCG AAAGGTCTTA CCGGAAAACT CGACGCAAGA AAAATCAGAG AGATCCTCAT 8880

AAAGGCCAAG AAGGGCGGAA AGTCCAAATT GTAAAATGTA ACTGTATTCA GCGATGACGA 8940

AATTCTTAGC TATTGTAATG ACTCTAGAGG ATCTTTGTGA AGGAACCTTA CTTCTGTGGT 9000

GTGACATAAT TGGACAAACT ACCTACAGAG ATTTAAAGCT CTAAGGTAAA TATAAAATTT 9060

TTAAGTGTAT AATGTGTTAA ACTACTGATT CTAATTGTTT GTGTATTTTA GATTCCAACC 9120

TATGGAACTG ATGAATGGGA GCAGTGGTGG AATGCCTTTA ATGAGGAAAA CCTGTTTTGC 9180

TCAGAAGAAA TGCCATCTAG TGATGATGAG GCTACTGCTG ACTCTCAACA TTCTACTCCT 9240

CCAAAAAAGA AGAGAAAGGT AGAAGACCCC AAGGACTTTC CTTCAGAATT GCTAAGTTTT 9300

TTGAGTCATG CTGTGTTTAG TAATAGAACT CTTGCTTGCT TTGCTATTTA CACCACAAAG 9360

GAAAAAGCTG CACTGCTATA CAAGAAAATT ATGGAAAAAT ATTCTGTAAC CTTTATAAGT 9420

AGGCATAACA GTTATAATCA TAACATACTG TTTTTTCTTA CTCCACACAG GCATAGAGTG 9480

TCTGCTATTA ATAACTATGC TCAAAAATTG TGTACCTTTA GCTTTTTAAT TTGTAAAGGG 9540

GTTAATAAGG AATATTTGAT GTATAGTGCC TTGACTAGAG ATCATAATCA GCCATACCAC 9600

ATTTGTAGAG GTTTTACTTG CTTTAAAAAA CCTCCCACAC CTCCCCCTGA ACCTGAAACA 9660

TAAAATGAAT GCAATTGTTG TTGTTAACTT GTTTATTGCA GCTTATAATG GTTACAAATA 9720

AAGCAATAGC ATCACAAATT TCACAAATAA AGCATTTTTT TCACTGCATT CTAGTTGTGG 9780

TTTGTCCAAA CTCATCAATG TATCTTATCA TGTCTGGATC CCCAGGAAGC TCCTCTGTGT 9840

CCTCATAAAC CCTAACCTCC TCTACTTGAG AGGACATTCC AATCATAGGC TGCCCATCCA 9900

CCCTCTGTGT CCTCCTGTTA ATTAGGTCAC TTAACAAAAA GGAAATTGGG TAGGGGTTTT 9960

TCACAGACCG CTTTCTAAGG GTAATTTTAA AATATCTGGG AAGTCCCTTC CACTGCTGTG 10020

TTCCAGAAGT GTTGGTAAAC AGCCCACAAA TGTCAACAGC AGAAACATAC AAGCTGTCAG 10080

CTTTGCACAA GGGCCCAACA CCCTGCTCAG CAAGAAGCAC TGTGGTTGCT GTGTTAGTAA 10140

TGTGCAAAAC AGGAGGCACA TTTTCCCCAC CTGTGTAGGT TCCAAAATAT CTAGTGTTTT 10200

CATTTTTACT TGGATCAGGA ACCCAGCACT CCACTGGATA AGCATTATCC TTATCCAAAA 10260

CAGCCTTGTG GTCAGTGTTC ATCTGCTGAC TGTCAACTGT AGCATTTTTT GGGGTTACAG 10320

TTTGAGCAGG ATATTTGGTC CTGTAGTTTG CTAACACACC CTGCAGCTCC AAAGGTTCCC 10380

CACCAACAGC AAAAAAATGA AAATTTGACC CTTGAATGGG TTTTCCAGCA CCATTTTCAT 10440

GAGTTTTTTG TGTCCCTGAA TGCAAGTTTA ACATAGCAGT TACCCCAATA ACCTCAGTTT 10500

TAACAGTAAC AGCTTCCCAC ATCAAAATAT TTCCACAGGT TAAGTCCTCA TTTAAATTAG 10560

GCAAAGGAA 10569

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10558 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATTCGA CCTCGAGGGA TCTTTGTGAA GGAACCTTAC TTCTGTGGTG 2160

TGACATAATT GGACAAACTA CCTACAGAGA TTTAAAGCTC TAAGGTAAAT ATAAAATTTT 2220

TAAGTGTATA ATGTGTTAAA CTACTGATTC TAATTGTTTG TGTATTTTAG ATTCCAACCT 2280

ATGGAACTGA TGAATGGGAG CAGTGGTGGA ATGCCTTTAA TGAGGAAAAC CTGTTTTGCT 2340

CAGAAGAAAT GCCATCTAGT GATGATGAGG CTACTGCTGA CTCTCAACAT TCTACTCCTC 2400

CAAAAAAGAA GAGAAAGGTA GAAGACCCCA AGGACTTTCC TTCAGAATTG CTAAGTTTTT 2460

TGAGTCATGC TGTGTTTAGT AATAGAACTC TTGCTTGCTT TGCTATTTAC ACCACAAAGG 2520

AAAAAGCTGC ACTGCTATAC AAGAAAATTA TGGAAAAATA TTCTGTAACC TTTATAAGTA 2580

GGCATAACAG TTATAATCAT AACATACTGT TTTTTCTTAC TCCACACAGG CATAGAGTGT 2640

CTGCTATTAA TAACTATGCT CAAAAATTGT GTACCTTTAG CTTTTTAATT TGTAAAGGGG 2700

TTAATAAGGA ATATTTGATG TATAGTGCCT TGACTAGAGA TCATAATCAG CCATACCACA 2760

TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC TCCCCCTGAA CCTGAAACAT 2820

AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG CTTATAATGG TTACAAATAA 2880

AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT 2940

TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC GGCTGTGGAA TGTGTGTCAG 3000

TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC 3060

AATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC CAGCAGGCAG AAGTATGCAA 3120

AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC 3180

CTAACTCCGC CCAGTTCCGC CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT 3240

GCAGAGGCCG AGGCCGCCTC GGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT 3300

GGAGGCCTAG GCTTTTGCAA AAAGCTTCAC GCTGCCGCAA GCACTCAGGG CGCAAGGGCT 3360

GCTAAAGGAA GCGGAACACG TAGAAAGCCA GTCCGCAGAA ACGGTGCTGA CCCCGGATGA 3420

ATGTCAGCTA CTGGGCTATC TGGACAAGGG AAAACGCAAG CGCAAAGAGA AAGCAGGTAG 3480

CTTGCAGTGG GCTTACATGG CGATAGCTAG ACTGGGCGGT TTTATGGACA GCAAGCGAAC 3540

CGGAATTGCC AGCTGGGGCG CCCTCTGGTA AGGTTGGGAA GCCCTGCAAA GTAAACTGGA 3600

TGGCTTTCTT GCCGCCAAGG ATCTGATGGC GCAGGGGATC AAGATCTGAT CAAGAGACAG 3660

GATGAGGATC GTTTCGCATG ATTGAACAAG ATGGATTGCA CGCAGGTTCT CCGGCCGCTT 3720

GGGTGGAGAG GCTATTCGGC TATGACTGGG CACAACAGAC AATCGGCTGC TCTGATGCCG 3780

CCGTGTTCCG GCTGTCAGCG CAGGGGCGCC CGGTTCTTTT TGTCAAGACC GACCTGTCCG 3840

GTGCCCTGAA TGAACTGCAG GACGAGGCAG CGCGGCTATC GTGGCTGGCC ACGACGGGCG 3900

TTCCTTGCGC AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG 3960

GCGAAGTGCC GGGGCAGGAT CTCCTGTCAT CTCACCTTGC TCCTGCCGAG AAAGTATCCA 4020

TCATGGCTGA TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC 4080

ACCAAGCGAA ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC 4140

AGGATGATCT GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTC GCCAGGCTCA 4200

AGGCGCGCAT GCCCGACGGC GAGGATCTCG TCGTGACCCA TGGCGATGCC TGCTTGCCGA 4260

ATATCATGGT GGAAAATGGC CGCTTTTCTG GATTCATCGA CTGTGGCCGG CTGGGTGTGG 4320

CGGACCGCTA TCAGGACATA GCGTTGGCTA CCCGTGATAT TGCTGAAGAG CTTGGCGGCG 4380

AATGGGCTGA CCGCTTCCTC GTGCTTTACG GTATCGCCGC TCCCGATTCG CAGCGCATCG 4440

CCTTCTATCG CCTTCTTGAC GAGTTCTTCT GAGCGGGACT CTGGGGTTCG AAATGACCGA 4500

CCAAGCGACG CCCAACCTGC CATCACGAGA TTTCGATTCC ACCGCCGCCT TCTATGAAAG 4560

GTTGGGCTTC GGAATCGTTT TCCGGGACGC CGGCTGGATG ATCCTCCAGC GCGGGGATCT 4620

CATGCTGGAG TTCTTCGCCC ACCCCGGGCT CGATCCCCTC GCGAGTTGGT TCAGCTGCTG 4680

CCTGAGGCTG GACGACCTCG CGGAGTTCTA CCGGCAGTGC AAATCCGTCG GCATCCAGGA 4740

AACCAGCAGC GGCTATCCGC GCATCCATGC CCCCGAACTG CAGGAGTGGG GAGGCACGAT 4800

GGCCGCTTTG GTCCCGGATC TTTGTGAAGG AACCTTACTT CTGTGGTGTG ACATAATTGG 4860

ACAAACTACC TACAGAGATT TAAAGCTCTA AGGTAAATAT AAAATTTTTA AGTGTATAAT 4920

GTGTTAAACT ACTGATTCTA ATTGTTTGTG TATTTTAGAT TCCAACCTAT GGAACTGATG 4980

AATGGGAGCA GTGGTGGAAT GCCTTTAATG AGGAAAACCT GTTTTGCTCA GAAGAAATGC 5040

CATCTAGTGA TGATGAGGCT ACTGCTGACT CTCAACATTC TACTCCTCCA AAAAAGAAGA 5100

GAAAGGTAGA AGACCCCAAG GACTTTCCTT CAGAATTGCT AAGTTTTTTG AGTCATGCTG 5160

TGTTTAGTAA TAGAACTCTT GCTTGCTTTG CTATTTACAC CACAAAGGAA AAAGCTGCAC 5220

TGCTATACAA GAAAATTATG GAAAAATATT CTGTAACCTT TATAAGTAGG CATAACAGTT 5280

ATAATCATAA CATACTGTTT TTTCTTACTC CACACAGGCA TAGAGTGTCT GCTATTAATA 5340

ACTATGCTCA AAAATTGTGT ACCTTTAGCT TTTTAATTTG TAAAGGGGTT AATAAGGAAT 5400

ATTTGATGTA TAGTGCCTTG ACTAGAGATC ATAATCAGCC ATACCACATT TGTAGAGGTT 5460

TTACTTGCTT TAAAAAACCT CCCACACCTC CCCCTGAACC TGAAACATAA AATGAATGCA 5520

ATTGTTGTTG TTAACTTGTT TATTGCAGCT TATAATGGTT ACAAATAAAG CAATAGCATC 5580

ACAAATTTCA CAAATAAAGC ATTTTTTTCA CTGCATTCTA GTTGTGGTTT GTCCAAACTC 5640

ATCAATGTAT CTTATCATGT CTGGATCCCC AGGAAGCTCC TCTGTGTCCT CATAAACCCT 5700

AACCTCCTCT ACTTGAGAGG ACATTCCAAT CATAGGCTGC CCATCCACCC TCTGTGTCCT 5760

CCTGTTAATT AGGTCACTTA ACAAAAAGGA AATTGGGTAG GGGTTTTTCA CAGACCGCTT 5820

TCTAAGGGTA ATTTTAAAAT ATCTGGGAAG TCCCTTCCAC TGCTGTGTTC CAGAAGTGTT 5880

GGTAAACAGC CCACAAATGT CAACAGCAGA AACATACAAG CTGTCAGCTT TGCACAAGGG 5940

CCCAACACCC TGCTCATCAA GAAGCACTGT GGTTGCTGTG TTAGTAATGT GCAAAACAGG 6000

AGGCACATTT TCCCCACCTG TGTAGGTTCC AAAATATCTA GTGTTTTCAT TTTTACTTGG 6060

ATCAGGAACC CAGCACTCCA CTGGATAAGC ATTATCCTTA TCCAAAACAG CCTTGTGGTC 6120

AGTGTTCATC TGCTGACTGT CAACTGTAGC ATTTTTTGGG GTTACAGTTT GAGCAGGATA 6180

TTTGGTCCTG TAGTTTGCTA ACACACCCTG CAGCTCCAAA GGTTCCCCAC CAACAGCAAA 6240

AAAATGAAAA TTTGACCCTT GAATGGGTTT TCCAGCACCA TTTTCATGAG TTTTTTGTGT 6300

CCCTGAATGC AAGTTTAACA TAGCAGTTAC CCCAATAACC TCAGTTTTAA CAGTAACAGC 6360

TTCCCACATC AAAATATTTC CACAGGTTAA GTCCTCATTT AAATTAGGCA AAGGAATTAT 6420

ACACTCCGCT ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG 6480

CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG 6540

TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGGCAGC 6600

GGATCATAAT CAGCCATACC ACATTTGTAG AGGTTTTACT TGCTTTAAAA AACCTCCCAC 6660

ACCTCCCCCT GAACCTGAAA CATAAAATGA ATGCAATTGT TGTTGTTAAC TTGTTTATTG 6720

CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT 6780

TTTCACTGCA TTCTAGTTGT GGTTTGTCCA AACTCATCAA TGTATCTTAT CATGTCTGGA 6840

TCATAATCAG CCATACCACA TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC 6900

TCCCCCTGAA CCTGAAACAT AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG 6960

CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT 7020

CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC 7080

CAGCCAGACA AGGTTGTTGA CACAAGACCC ACATCTGGTA TAAAAGGAGG CAGTGGCCCA 7140

CAGAGGAGCA CAGCTGTGTT TGGCTGCAGG GCCAAGAGCG CTGTCAAGAA GACCCACACG 7200

CCCCCCTCCA GCAGCTGAAT TCCAGCTGGC ATTCCGGTAC TGTTGGTAAA ATGGAAGACG 7260

CCAAAAACAT AAAGAAAGGC CCGGCGCCAT TCTATCCTCT AGAGGATGGA ACCGCTGGAG 7320

AGCAACTGCA TAAGGCTATG AAGAGATACG CCCTGGTTCC TGGAACAATT GCTTTTACAG 7380

ATGCACATAT CGAGGTGAAC ATCACGTACG CGGAATACTT CGAAATGTCC GTTCGGTTGG 7440

CAGAAGCTAT GAAACGATAT GGGCTGAATA CAAATCACAG AATCGTCGTA TGCAGTGAAA 7500

ACTCTCTTCA ATTCTTTATG CCGGTGTTGG GCGCGTTATT TATCGGAGTT GCAGTTGCGC 7560

CCGCGAACGA CATTTATAAT GAACGTGAAT TGCTCAACAG TATGAACATT TCGCAGCCTA 7620

CCGTAGTGTT TGTTTCCAAA AAGGGGTTGC AAAAAATTTT GAACGTGCAA AAAAAATTAC 7680

CAATAATCCA GAAAATTATT ATCATGGATT CTAAAACGGA TTACCAGGGA TTTCAGTCGA 7740

TGTACACGTT CGTCACATCT CATCTACCTC CCGGTTTTAA TGAATACGAT TTTGTACCAG 7800

AGTCCTTTGA TCGTGACAAA ACAATTGCAC TGATAATGAA TTCCTCTGGA TCTACTGGGT 7860

TACCTAAGGG TGTGGCCCTT CCGCATAGAA CTGCCTGCGT CAGATTCTCG CATGCCAGAG 7920

ATCCTATTTT TGGCAATCAA ATCATTCCGG ATACTGCGAT TTTAAGTGTT GTTCCATTCC 7980

ATCACGGTTT TGGAATGTTT ACTACACTCG GATATTTGAT ATGTGGATTT CGAGTCGTCT 8040

TAATGTATAG ATTTGAAGAA GAGCTGTTTT TACGATCCCT TCAGGATTAC AAAATTCAAA 8100

GTGCGTTGCT AGTACCAACC CTATTTTCAT TCTTCGCCAA AAGCACTCTG ATTGACAAAT 8160

ACGATTTATC TAATTTACAC GAAATTGCTT CTGGGGGCGC ACCTCTTTCG AAAGAAGTCG 8220

GGGAAGCGGT TGCAAAACGC TTCCATCTTC CAGGGATACG ACAAGGATAT GGGCTCACTG 8280

AGACTACATC AGCTATTCTG ATTACACCCG AGGGGGATGA TAAACCGGGC GCGGTCGGTA 8340

AAGTTGTTCC ATTTTTTGAA GCGAAGGTTG TGGATCTGGA TACCGGGAAA ACGCTGGGCG 8400

TTAATCAGAG AGGCGAATTA TGTGTCAGAG GACCTATGAT TATGTCCGGT TATGTAAACA 8460

ATCCGGAAGC GACCAACGCC TTGATTGACA AGGATGGATG GCTACATTCT GGAGACATAG 8520

CTTACTGGGA CGAAGACGAA CACTTCTTCA TAGTTGACCG CTTGAAGTCT TTAATTAAAT 8580

ACAAAGGATA TCAGGTGGCC CCCGCTGAAT TGGAATCGAT ATTGTTACAA CACCCCAACA 8640

TCTTCGACGC GGGCGTGGCA GGTCTTCCCG ACGATGACGC CGGTGAACTT CCCGCCGCCG 8700

TTGTTGTTTT GGAGCACGGA AAGACGATGA CGGAAAAAGA GATCGTGGAT TACGTCGCCA 8760

GTCAAGTAAC AACCGCGAAA AAGTTGCGCG GAGGAGTTGT GTTTGTGGAC GAAGTACCGA 8820

AAGGTCTTAC CGGAAAACTC GACGCAAGAA AAATCAGAGA GATCCTCATA AAGGCCAAGA 8880

AGGGCGGAAA GTCCAAATTG TAAAATGTAA CTGTATTCAG CGATGACGAA ATTCTTAGCT 8940

ATTGTAATGA CTCTAGAGGA TCTTTGTGAA GGAACCTTAC TTCTGTGGTG TGACATAATT 9000

GGACAAACTA CCTACAGAGA TTTAAAGCTC TAAGGTAAAT ATAAAATTTT TAAGTGTATA 9060

ATGTGTTAAA CTACTGATTC TAATTGTTTG TGTATTTTAG ATTCCAACCT ATGGAACTGA 9120

TGAATGGGAG CAGTGGTGGA ATGCCTTTAA TGAGGAAAAC CTGTTTTGCT CAGAAGAAAT 9180

GCCATCTAGT GATGATGAGG CTACTGCTGA CTCTCAACAT TCTACTCCTC CAAAAAAGAA 9240

GAGAAAGGTA GAAGACCCCA AGGACTTTCC TTCAGAATTG CTAAGTTTTT TGAGTCATGC 9300

TGTGTTTAGT AATAGAACTC TTGCTTGCTT TGCTATTTAC ACCACAAAGG AAAAAGCTGC 9360

ACTGCTATAC AAGAAAATTA TGGAAAAATA TTCTGTAACC TTTATAAGTA GGCATAACAG 9420

TTATAATCAT AACATACTGT TTTTTCTTAC TCCACACAGG CATAGAGTGT CTGCTATTAA 9480

TAACTATGCT CAAAAATTGT GTACCTTTAG CTTTTTAATT TGTAAAGGGG TTAATAAGGA 9540

ATATTTGATG TATAGTGCCT TGACTAGAGA TCATAATCAG CCATACCACA TTTGTAGAGG 9600

TTTTACTTGC TTTAAAAAAC CTCCCACACC TCCCCCTGAA CCTGAAACAT AAAATGAATG 9660

CAATTGTTGT TGTTAACTTG TTTATTGCAG CTTATAATGG TTACAAATAA AGCAATAGCA 9720

TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT TTGTCCAAAC 9780

TCATCAATGT ATCTTATCAT GTCTGGATCC CCAGGAAGCT CCTCTGTGTC CTCATAAACC 9840

CTAACCTCCT CTACTTGAGA GGACATTCCA ATCATAGGCT GCCCATCCAC CCTCTGTGTC 9900

CTCCTGTTAA TTAGGTCACT TAACAAAAAG GAAATTGGGT AGGGGTTTTT CACAGACCGC 9960

TTTCTAAGGG TAATTTTAAA ATATCTGGGA AGTCCCTTCC ACTGCTGTGT TCCAGAAGTG 10020

TTGGTAAACA GCCCACAAAT GTCAACAGCA GAAACATACA AGCTGTCAGC TTTGCACAAG 10080

GGCCCAACAC CCTGCTCAGC AAGAAGCACT GTGGTTGCTG TGTTAGTAAT GTGCAAAACA 10140

GGAGGCACAT TTTCCCCACC TGTGTAGGTT CCAAAATATC TAGTGTTTTC ATTTTTACTT 10200

GGATCAGGAA CCCAGCACTC CACTGGATAA GCATTATCCT TATCCAAAAC AGCCTTGTGG 10260

TCAGTGTTCA TCTGCTGACT GTCAACTGTA GCATTTTTTG GGGTTACAGT TTGAGCAGGA 10320

TATTTGGTCC TGTAGTTTGC TAACACACCC TGCAGCTCCA AAGGTTCCCC ACCAACAGCA 10380

AAAAAATGAA AATTTGACCC TTGAATGGGT TTTCCAGCAC CATTTTCATG AGTTTTTTGT 10440

GTCCCTGAAT GCAAGTTTAA CATAGCAGTT ACCCCAATAA CCTCAGTTTT AACAGTAACA 10500

GCTTCCCACA TCAAAATATT TCCACAGGTT AAGTCCTCAT TTAAATTAGG CAAAGGAA 10558

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6245 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATACAC TCCGCTATCG CTACGTGACT GGGTCATGGC TGCGCCCCGA 2160

CACCCGCCAA CACCCGCTGA CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC 2220

AGACAAGCTG TGACCGTCTC CGGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG 2280

AAACGCGCGA GGCAGCGGAT CATAATCAGC CATACCACAT TTGTAGAGGT TTTACTTGCT 2340

TTAAAAAACC TCCCACACCT CCCCCTGAAC CTGAAACATA AAATGAATGC AATTGTTGTT 2400

GTTAACTTGT TTATTGCAGC TTATAATGGT TACAAATAAA GCAATAGCAT CACAAATTTC 2460

ACAAATAAAG CATTTTTTTC ACTGCATTCT AGTTGTGGTT TGTCCAAACT CATCAATGTA 2520

TCTTATCATG TCTGGATCAT AATCAGCCAT ACCACATTTG TAGAGGTTTT ACTTGCTTTA 2580

AAAAACCTCC CACACCTCCC CCTGAACCTG AAACATAAAA TGAATGCAAT TGTTGTTGTT 2640

AACTTGTTTA TTGCAGCTTA TAATGGTTAC AAATAAAGCA ATAGCATCAC AAATTTCACA 2700

AATAAAGCAT TTTTTTCACT GCATTCTAGT TGTGGTTTGT CCAAACTCAT CAATGTATCT 2760

TATCATGTCT GGATCCCAAG TTCATCTATT TCCTCCCACA TCTGGTATAA AAGGAGGCAG 2820

TGGCCCACAG AGGAGCACAG CTGTGTTTGG CTGCAGGGCC AAGAGCGCTG TCAAGAAGAC 2880

CCACACGCCC CCCTCCAGCA GCTGAATTCC AGCTGGCATT CCGGTACTGT TGGTAAAATG 2940

GAAGACGCCA AAAACATAAA GAAAGGCCCG GCGCCATTCT ATCCTCTAGA GGATGGAACC 3000

GCTGGAGAGC AACTGCATAA GGCTATGAAG AGATACGCCC TGGTTCCTGG AACAATTGCT 3060

TTTACAGATG CACATATCGA GGTGAACATC ACGTACGCGG AATACTTCGA AATGTCCGTT 3120

CGGTTGGCAG AAGCTATGAA ACGATATGGG CTGAATACAA ATCACAGAAT CGTCGTATGC 3180

AGTGAAAACT CTCTTCAATT CTTTATGCCG GTGTTGGGCG CGTTATTTAT CGGAGTTGCA 3240

GTTGCGCCCG CGAACGACAT TTATAATGAA CGTGAATTGC TCAACAGTAT GAACATTTCG 3300

CAGCCTACCG TAGTGTTTGT TTCCAAAAAG GGGTTGCAAA AAATTTTGAA CGTGCAAAAA 3360

AAATTACCAA TAATCCAGAA AATTATTATC ATGGATTCTA AAACGGATTA CCAGGGATTT 3420

CAGTCGATGT ACACGTTCGT CACATCTCAT CTACCTCCCG GTTTTAATGA ATACGATTTT 3480

GTACCAGAGT CCTTTGATCG TGACAAAACA ATTGCACTGA TAATGAATTC CTCTGGATCT 3540

ACTGGGTTAC CTAAGGGTGT GGCCCTTCCG CATAGAACTG CCTGCGTCAG ATTCTCGCAT 3600

GCCAGAGATC CTATTTTTGG CAATCAAATC ATTCCGGATA CTGCGATTTT AAGTGTTGTT 3660

CCATTCCATC ACGGTTTTGG AATGTTTACT ACACTCGGAT ATTTGATATG TGGATTTCGA 3720

GTCGTCTTAA TGTATAGATT TGAAGAAGAG CTGTTTTTAC GATCCCTTCA GGATTACAAA 3780

ATTCAAAGTG CGTTGCTAGT ACCAACCCTA TTTTCATTCT TCGCCAAAAG CACTCTGATT 3840

GACAAATACG ATTTATCTAA TTTACACGAA ATTGCTTCTG GGGGCGCACC TCTTTCGAAA 3900

GAAGTCGGGG AAGCGGTTGC AAAACGCTTC CATCTTCCAG GGATACGACA AGGATATGGG 3960

CTCACTGAGA CTACATCAGC TATTCTGATT ACACCCGAGG GGGATGATAA ACCGGGCGCG 4020

GTCGGTAAAG TTGTTCCATT TTTTGAAGCG AAGGTTGTGG ATCTGGATAC CGGGAAAACG 4080

CTGGGCGTTA ATCAGAGAGG CGAATTATGT GTCAGAGGAC CTATGATTAT GTCCGGTTAT 4140

GTAAACAATC CGGAAGCGAC CAACGCCTTG ATTGACAAGG ATGGATGGCT ACATTCTGGA 4200

GACATAGCTT ACTGGGACGA AGACGAACAC TTCTTCATAG TTGACCGCTT GAAGTCTTTA 4260

ATTAAATACA AAGGATATCA GGTGGCCCCC GCTGAATTGG AATCGATATT GTTACAACAC 4320

CCCAACATCT TCGACGCGGG CGTGGCAGGT CTTCCCGACG ATGACGCCGG TGAACTTCCC 4380

GCCGCCGTTG TTGTTTTGGA GCACGGAAAG ACGATGACGG AAAAAGAGAT CGTGGATTAC 4440

GTCGCCAGTC AAGTAACAAC CGCGAAAAAG TTGCGCGGAG GAGTTGTGTT TGTGGACGAA 4500

GTACCGAAAG GTCTTACCGG AAAACTCGAC GCAAGAAAAA TCAGAGAGAT CCTCATAAAG 4560

GCCAAGAAGG GCGGAAAGTC CAAATTGTAA AATGTAACTG TATTCAGCGA TGACGAAATT 4620

CTTAGCTATT GTAATGACTC TAGAGGATCT TTGTGAAGGA ACCTTACTTC TGTGGTGTGA 4680

CATAATTGGA CAAACTACCT ACAGAGATTT AAAGCTCTAA GGTAAATATA AAATTTTTAA 4740

GTGTATAATG TGTTAAACTA CTGATTCTAA TTGTTTGTGT ATTTTAGATT CCAACCTATG 4800

GAACTGATGA ATGGGAGCAG TGGTGGAATG CCTTTAATGA GGAAAACCTG TTTTGCTCAG 4860

AAGAAATGCC ATCTAGTGAT GATGAGGCTA CTGCTGACTC TCAACATTCT ACTCCTCCAA 4920

AAAAGAAGAG AAAGGTAGAA GACCCCAAGG ACTTTCCTTC AGAATTGCTA AGTTTTTTGA 4980

GTCATGCTGT GTTTAGTAAT AGAACTCTTG CTTGCTTTGC TATTTACACC ACAAAGGAAA 5040

AAGCTGCACT GCTATACAAG AAAATTATGG AAAAATATTC TGTAACCTTT ATAAGTAGGC 5100

ATAACAGTTA TAATCATAAC ATACTGTTTT TTCTTACTCC ACACAGGCAT AGAGTGTCTG 5160

CTATTAATAA CTATGCTCAA AAATTGTGTA CCTTTAGCTT TTTAATTTGT AAAGGGGTTA 5220

ATAAGGAATA TTTGATGTAT AGTGCCTTGA CTAGAGATCA TAATCAGCCA TACCACATTT 5280

GTAGAGGTTT TACTTGCTTT AAAAAACCTC CCACACCTCC CCCTGAACCT GAAACATAAA 5340

ATGAATGCAA TTGTTGTTGT TAACTTGTTT ATTGCAGCTT ATAATGGTTA CAAATAAAGC 5400

AATAGCATCA CAAATTTCAC AAATAAAGCA TTTTTTTCAC TGCATTCTAG TTGTGGTTTG 5460

TCCAAACTCA TCAATGTATC TTATCATGTC TGGATCCCCA GGAAGCTCCT CTGTGTCCTC 5520

ATAAACCCTA ACCTCCTCTA CTTGAGAGGA CATTCCAATC ATAGGCTGCC CATCCACCCT 5580

CTGTGTCCTC CTGTTAATTA GGTCACTTAA CAAAAAGGAA ATTGGGTAGG GGTTTTTCAC 5640

AGACCGCTTT CTAAGGGTAA TTTTAAAATA TCTGGGAAGT CCCTTCCACT GCTGTGTTCC 5700

AGAAGTGTTG GTAAACAGCC CACAAATGTC AACAGCAGAA ACATACAAGC TGTCAGCTTT 5760

GCACAAGGGC CCAACACCCT GCTCAGCAAG AAGCACTGTG GTTGCTGTGT TAGTAATGTG 5820

CAAAACAGGA GGCACATTTT CCCCACCTGT GTAGGTTCCA AAATATCTAG TGTTTTCATT 5880

TTTACTTGGA TCAGGAACCC AGCACTCCAC TGGATAAGCA TTATCCTTAT CCAAAACAGC 5940

CTTGTGGTCA GTGTTCATCT GCTGACTGTC AACTGTAGCA TTTTTTGGGG TTACAGTTTG 6000

AGCAGGATAT TTGGTCCTGT AGTTTGCTAA CACACCCTGC AGCTCCAAAG GTTCCCCACC 6060

AACAGCAAAA AAATGAAAAT TTGACCCTTG AATGGGTTTT CCAGCACCAT TTTCATGAGT 6120

TTTTTGTGTC CCTGAATGCA AGTTTAACAT AGCAGTTACC CCAATAACCT CAGTTTTAAC 6180

AGTAACAGCT TCCCACATCA AAATATTTCC ACAGGTTAAG TCCTCATTTA AATTAGGCAA 6240

AGGAA 6245

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6254 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATACAC TCCGCTATCG CTACGTGACT GGGTCATGGC TGCGCCCCGA 2160

CACCCGCCAA CACCCGCTGA CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC 2220

AGACAAGCTG TGACCGTCTC CGGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG 2280

AAACGCGCGA GGCAGCGGAT CATAATCAGC CATACCACAT TTGTAGAGGT TTTACTTGCT 2340

TTAAAAAACC TCCCACACCT CCCCCTGAAC CTGAAACATA AAATGAATGC AATTGTTGTT 2400

GTTAACTTGT TTATTGCAGC TTATAATGGT TACAAATAAA GCAATAGCAT CACAAATTTC 2460

ACAAATAAAG CATTTTTTTC ACTGCATTCT AGTTGTGGTT TGTCCAAACT CATCAATGTA 2520

TCTTATCATG TCTGGATCAT AATCAGCCAT ACCACATTTG TAGAGGTTTT ACTTGCTTTA 2580

AAAAACCTCC CACACCTCCC CCTGAACCTG AAACATAAAA TGAATGCAAT TGTTGTTGTT 2640

AACTTGTTTA TTGCAGCTTA TAATGGTTAC AAATAAAGCA ATAGCATCAC AAATTTCACA 2700

AATAAAGCAT TTTTTTCACT GCATTCTAGT TGTGGTTTGT CCAAACTCAT CAATGTATCT 2760

TATCATGTCT GGATCCCAGT GGGGAGTCAG CCGTGTATCA TCGCCCACAT CTGGTATAAA 2820

AGGAGGCAGT GGCCCACAGA GGAGCACAGC TGTGTTTGGC TGCAGGGCCA AGAGCGCTGT 2880

CAAGAAGACC CACACGCCCC CCTCCAGCAG CTGAATTCCA GCTGGCATTC CGGTACTGTT 2940

GGTAAAATGG AAGACGCCAA AAACATAAAG AAAGGCCCGG CGCCATTCTA TCCTCTAGAG 3000

GATGGAACCG CTGGAGAGCA ACTGCATAAG GCTATGAAGA GATACGCCCT GGTTCCTGGA 3060

ACAATTGCTT TTACAGATGC ACATATCGAG GTGAACATCA CGTACGCGGA ATACTTCGAA 3120

ATGTCCGTTC GGTTGGCAGA AGCTATGAAA CGATATGGGC TGAATACAAA TCACAGAATC 3180

GTCGTATGCA GTGAAAACTC TCTTCAATTC TTTATGCCGG TGTTGGGCGC GTTATTTATC 3240

GGAGTTGCAG TTGCGCCCGC GAACGACATT TATAATGAAC GTGAATTGCT CAACAGTATG 3300

AACATTTCGC AGCCTACCGT AGTGTTTGTT TCCAAAAAGG GGTTGCAAAA AATTTTGAAC 3360

GTGCAAAAAA AATTACCAAT AATCCAGAAA ATTATTATCA TGGATTCTAA AACGGATTAC 3420

CAGGGATTTC AGTCGATGTA CACGTTCGTC ACATCTCATC TACCTCCCGG TTTTAATGAA 3480

TACGATTTTG TACCAGAGTC CTTTGATCGT GACAAAACAA TTGCACTGAT AATGAATTCC 3540

TCTGGATCTA CTGGGTTACC TAAGGGTGTG GCCCTTCCGC ATAGAACTGC CTGCGTCAGA 3600

TTCTCGCATG CCAGAGATCC TATTTTTGGC AATCAAATCA TTCCGGATAC TGCGATTTTA 3660

AGTGTTGTTC CATTCCATCA CGGTTTTGGA ATGTTTACTA CACTCGGATA TTTGATATGT 3720

GGATTTCGAG TCGTCTTAAT GTATAGATTT GAAGAAGAGC TGTTTTTACG ATCCCTTCAG 3780

GATTACAAAA TTCAAAGTGC GTTGCTAGTA CCAACCCTAT TTTCATTCTT CGCCAAAAGC 3840

ACTCTGATTG ACAAATACGA TTTATCTAAT TTACACGAAA TTGCTTCTGG GGGCGCACCT 3900

CTTTCGAAAG AAGTCGGGGA AGCGGTTGCA AAACGCTTCC ATCTTCCAGG GATACGACAA 3960

GGATATGGGC TCACTGAGAC TACATCAGCT ATTCTGATTA CACCCGAGGG GGATGATAAA 4020

CCGGGCGCGG TCGGTAAAGT TGTTCCATTT TTTGAAGCGA AGGTTGTGGA TCTGGATACC 4080

GGGAAAACGC TGGGCGTTAA TCAGAGAGGC GAATTATGTG TCAGAGGACC TATGATTATG 4140

TCCGGTTATG TAAACAATCC GGAAGCGACC AACGCCTTGA TTGACAAGGA TGGATGGCTA 4200

CATTCTGGAG ACATAGCTTA CTGGGACGAA GACGAACACT TCTTCATAGT TGACCGCTTG 4260

AAGTCTTTAA TTAAATACAA AGGATATCAG GTGGCCCCCG CTGAATTGGA ATCGATATTG 4320

TTACAACACC CCAACATCTT CGACGCGGGC GTGGCAGGTC TTCCCGACGA TGACGCCGGT 4380

GAACTTCCCG CCGCCGTTGT TGTTTTGGAG CACGGAAAGA CGATGACGGA AAAAGAGATC 4440

GTGGATTACG TCGCCAGTCA AGTAACAACC GCGAAAAAGT TGCGCGGAGG AGTTGTGTTT 4500

GTGGACGAAG TACCGAAAGG TCTTACCGGA AAACTCGACG CAAGAAAAAT CAGAGAGATC 4560

CTCATAAAGG CCAAGAAGGG CGGAAAGTCC AAATTGTAAA ATGTAACTGT ATTCAGCGAT 4620

GACGAAATTC TTAGCTATTG TAATGACTCT AGAGGATCTT TGTGAAGGAA CCTTACTTCT 4680

GTGGTGTGAC ATAATTGGAC AAACTACCTA CAGAGATTTA AAGCTCTAAG GTAAATATAA 4740

AATTTTTAAG TGTATAATGT GTTAAACTAC TGATTCTAAT TGTTTGTGTA TTTTAGATTC 4800

CAACCTATGG AACTGATGAA TGGGAGCAGT GGTGGAATGC CTTTAATGAG GAAAACCTGT 4860

TTTGCTCAGA AGAAATGCCA TCTAGTGATG ATGAGGCTAC TGCTGACTCT CAACATTCTA 4920

CTCCTCCAAA AAAGAAGAGA AAGGTAGAAG ACCCCAAGGA CTTTCCTTCA GAATTGCTAA 4980

GTTTTTTGAG TCATGCTGTG TTTAGTAATA GAACTCTTGC TTGCTTTGCT ATTTACACCA 5040

CAAAGGAAAA AGCTGCACTG CTATACAAGA AAATTATGGA AAAATATTCT GTAACCTTTA 5100

TAAGTAGGCA TAACAGTTAT AATCATAACA TACTGTTTTT TCTTACTCCA CACAGGCATA 5160

GAGTGTCTGC TATTAATAAC TATGCTCAAA AATTGTGTAC CTTTAGCTTT TTAATTTGTA 5220

AAGGGGTTAA TAAGGAATAT TTGATGTATA GTGCCTTGAC TAGAGATCAT AATCAGCCAT 5280

ACCACATTTG TAGAGGTTTT ACTTGCTTTA AAAAACCTCC CACACCTCCC CCTGAACCTG 5340

AAACATAAAA TGAATGCAAT TGTTGTTGTT AACTTGTTTA TTGCAGCTTA TAATGGTTAC 5400

AAATAAAGCA ATAGCATCAC AAATTTCACA AATAAAGCAT TTTTTTCACT GCATTCTAGT 5460

TGTGGTTTGT CCAAACTCAT CAATGTATCT TATCATGTCT GGATCCCCAG GAAGCTCCTC 5520

TGTGTCCTCA TAAACCCTAA CCTCCTCTAC TTGAGAGGAC ATTCCAATCA TAGGCTGCCC 5580

ATCCACCCTC TGTGTCCTCC TGTTAATTAG GTCACTTAAC AAAAAGGAAA TTGGGTAGGG 5640

GTTTTTCACA GACCGCTTTC TAAGGGTAAT TTTAAAATAT CTGGGAAGTC CCTTCCACTG 5700

CTGTGTTCCA GAAGTGTTGG TAAACAGCCC ACAAATGTCA ACAGCAGAAA CATACAAGCT 5760

GTCAGCTTTG CACAAGGGCC CAACACCCTG CTCAGCAAGA AGCACTGTGG TTGCTGTGTT 5820

AGTAATGTGC AAAACAGGAG GCACATTTTC CCCACCTGTG TAGGTTCCAA AATATCTAGT 5880

GTTTTCATTT TTACTTGGAT CAGGAACCCA GCACTCCACT GGATAAGCAT TATCCTTATC 5940

CAAAACAGCC TTGTGGTCAG TGTTCATCTG CTGACTGTCA ACTGTAGCAT TTTTTGGGGT 6000

TACAGTTTGA GCAGGATATT TGGTCCTGTA GTTTGCTAAC ACACCCTGCA GCTCCAAAGG 6060

TTCCCCACCA ACAGCAAAAA AATGAAAATT TGACCCTTGA ATGGGTTTTC CAGCACCATT 6120

TTCATGAGTT TTTTGTGTCC CTGAATGCAA GTTTAACATA GCAGTTACCC CAATAACCTC 6180

AGTTTTAACA GTAACAGCTT CCCACATCAA AATATTTCCA CAGGTTAAGT CCTCATTTAA 6240

ATTAGGCAAA GGAA 6254

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6265 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATACAC TCCGCTATCG CTACGTGACT GGGTCATGGC TGCGCCCCGA 2160

CACCCGCCAA CACCCGCTGA CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC 2220

AGACAAGCTG TGACCGTCTC CGGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG 2280

AAACGCGCGA GGCAGCGGAT CATAATCAGC CATACCACAT TTGTAGAGGT TTTACTTGCT 2340

TTAAAAAACC TCCCACACCT CCCCCTGAAC CTGAAACATA AAATGAATGC AATTGTTGTT 2400

GTTAACTTGT TTATTGCAGC TTATAATGGT TACAAATAAA GCAATAGCAT CACAAATTTC 2460

ACAAATAAAG CATTTTTTTC ACTGCATTCT AGTTGTGGTT TGTCCAAACT CATCAATGTA 2520

TCTTATCATG TCTGGATCAT AATCAGCCAT ACCACATTTG TAGAGGTTTT ACTTGCTTTA 2580

AAAAACCTCC CACACCTCCC CCTGAACCTG AAACATAAAA TGAATGCAAT TGTTGTTGTT 2640

AACTTGTTTA TTGCAGCTTA TAATGGTTAC AAATAAAGCA ATAGCATCAC AAATTTCACA 2700

AATAAAGCAT TTTTTTCACT GCATTCTAGT TGTGGTTTGT CCAAACTCAT CAATGTATCT 2760

TATCATGTCT GGATCCCACT CCAACCTCAG CCAGACAAGG TTGTTGACAC AAGACCCACA 2820

TCTGGTATAA AAGGAGGCAG TGGCCCACAG AGGAGCACAG CTGTGTTTGG CTGCAGGGCC 2880

AAGAGCGCTG TCAAGAAGAC CCACACGCCC CCCTCCAGCA GCTGAATTCC AGCTGGCATT 2940

CCGGTACTGT TGGTAAAATG GAAGACGCCA AAAACATAAA GAAAGGCCCG GCGCCATTCT 3000

ATCCTCTAGA GGATGGAACC GCTGGAGAGC AACTGCATAA GGCTATGAAG AGATACGCCC 3060

TGGTTCCTGG AACAATTGCT TTTACAGATG CACATATCGA GGTGAACATC ACGTACGCGG 3120

AATACTTCGA AATGTCCGTT CGGTTGGCAG AAGCTATGAA ACGATATGGG CTGAATACAA 3180

ATCACAGAAT CGTCGTATGC AGTGAAAACT CTCTTCAATT CTTTATGCCG GTGTTGGGCG 3240

CGTTATTTAT CGGAGTTGCA GTTGCGCCCG CGAACGACAT TTATAATGAA CGTGAATTGC 3300

TCAACAGTAT GAACATTTCG CAGCCTACCG TAGTGTTTGT TTCCAAAAAG GGGTTGCAAA 3360

AAATTTTGAA CGTGCAAAAA AAATTACCAA TAATCCAGAA AATTATTATC ATGGATTCTA 3420

AAACGGATTA CCAGGGATTT CAGTCGATGT ACACGTTCGT CACATCTCAT CTACCTCCCG 3480

GTTTTAATGA ATACGATTTT GTACCAGAGT CCTTTGATCG TGACAAAACA ATTGCACTGA 3540

TAATGAATTC CTCTGGATCT ACTGGGTTAC CTAAGGGTGT GGCCCTTCCG CATAGAACTG 3600

CCTGCGTCAG ATTCTCGCAT GCCAGAGATC CTATTTTTGG CAATCAAATC ATTCCGGATA 3660

CTGCGATTTT AAGTGTTGTT CCATTCCATC ACGGTTTTGG AATGTTTACT ACACTCGGAT 3720

ATTTGATATG TGGATTTCGA GTCGTCTTAA TGTATAGATT TGAAGAAGAG CTGTTTTTAC 3780

GATCCCTTCA GGATTACAAA ATTCAAAGTG CGTTGCTAGT ACCAACCCTA TTTTCATTCT 3840

TCGCCAAAAG CACTCTGATT GACAAATACG ATTTATCTAA TTTACACGAA ATTGCTTCTG 3900

GGGGCGCACC TCTTTCGAAA GAAGTCGGGG AAGCGGTTGC AAAACGCTTC CATCTTCCAG 3960

GGATACGACA AGGATATGGG CTCACTGAGA CTACATCAGC TATTCTGATT ACACCCGAGG 4020

GGGATGATAA ACCGGGCGCG GTCGGTAAAG TTGTTCCATT TTTTGAAGCG AAGGTTGTGG 4080

ATCTGGATAC CGGGAAAACG CTGGGCGTTA ATCAGAGAGG CGAATTATGT GTCAGAGGAC 4140

CTATGATTAT GTCCGGTTAT GTAAACAATC CGGAAGCGAC CAACGCCTTG ATTGACAAGG 4200

ATGGATGGCT ACATTCTGGA GACATAGCTT ACTGGGACGA AGACGAACAC TTCTTCATAG 4260

TTGACCGCTT GAAGTCTTTA ATTAAATACA AAGGATATCA GGTGGCCCCC GCTGAATTGG 4320

AATCGATATT GTTACAACAC CCCAACATCT TCGACGCGGG CGTGGCAGGT CTTCCCGACG 4380

ATGACGCCGG TGAACTTCCC GCCGCCGTTG TTGTTTTGGA GCACGGAAAG ACGATGACGG 4440

AAAAAGAGAT CGTGGATTAC GTCGCCAGTC AAGTAACAAC CGCGAAAAAG TTGCGCGGAG 4500

GAGTTGTGTT TGTGGACGAA GTACCGAAAG GTCTTACCGG AAAACTCGAC GCAAGAAAAA 4560

TCAGAGAGAT CCTCATAAAG GCCAAGAAGG GCGGAAAGTC CAAATTGTAA AATGTAACTG 4620

TATTCAGCGA TGACGAAATT CTTAGCTATT GTAATGACTC TAGAGGATCT TTGTGAAGGA 4680

ACCTTACTTC TGTGGTGTGA CATAATTGGA CAAACTACCT ACAGAGATTT AAAGCTCTAA 4740

GGTAAATATA AAATTTTTAA GTGTATAATG TGTTAAACTA CTGATTCTAA TTGTTTGTGT 4800

ATTTTAGATT CCAACCTATG GAACTGATGA ATGGGAGCAG TGGTGGAATG CCTTTAATGA 4860

GGAAAACCTG TTTTGCTCAG AAGAAATGCC ATCTAGTGAT GATGAGGCTA CTGCTGACTC 4920

TCAACATTCT ACTCCTCCAA AAAAGAAGAG AAAGGTAGAA GACCCCAAGG ACTTTCCTTC 4980

AGAATTGCTA AGTTTTTTGA GTCATGCTGT GTTTAGTAAT AGAACTCTTG CTTGCTTTGC 5040

TATTTACACC ACAAAGGAAA AAGCTGCACT GCTATACAAG AAAATTATGG AAAAATATTC 5100

TGTAACCTTT ATAAGTAGGC ATAACAGTTA TAATCATAAC ATACTGTTTT TTCTTACTCC 5160

ACACAGGCAT AGAGTGTCTG CTATTAATAA CTATGCTCAA AAATTGTGTA CCTTTAGCTT 5220

TTTAATTTGT AAAGGGGTTA ATAAGGAATA TTTGATGTAT AGTGCCTTGA CTAGAGATCA 5280

TAATCAGCCA TACCACATTT GTAGAGGTTT TACTTGCTTT AAAAAACCTC CCACACCTCC 5340

CCCTGAACCT GAAACATAAA ATGAATGCAA TTGTTGTTGT TAACTTGTTT ATTGCAGCTT 5400

ATAATGGTTA CAAATAAAGC AATAGCATCA CAAATTTCAC AAATAAAGCA TTTTTTTCAC 5460

TGCATTCTAG TTGTGGTTTG TCCAAACTCA TCAATGTATC TTATCATGTC TGGATCCCCA 5520

GGAAGCTCCT CTGTGTCCTC ATAAACCCTA ACCTCCTCTA CTTGAGAGGA CATTCCAATC 5580

ATAGGCTGCC CATCCACCCT CTGTGTCCTC CTGTTAATTA GGTCACTTAA CAAAAAGGAA 5640

ATTGGGTAGG GGTTTTTCAC AGACCGCTTT CTAAGGGTAA TTTTAAAATA TCTGGGAAGT 5700

CCCTTCCACT GCTGTGTTCC AGAAGTGTTG GTAAACAGCC CACAAATGTC AACAGCAGAA 5760

ACATACAAGC TGTCAGCTTT GCACAAGGGC CCAACACCCT GCTCAGCAAG AAGCACTGTG 5820

GTTGCTGTGT TAGTAATGTG CAAAACAGGA GGCACATTTT CCCCACCTGT GTAGGTTCCA 5880

AAATATCTAG TGTTTTCATT TTTACTTGGA TCAGGAACCC AGCACTCCAC TGGATAAGCA 5940

TTATCCTTAT CCAAAACAGC CTTGTGGTCA GTGTTCATCT GCTGACTGTC AACTGTAGCA 6000

TTTTTTGGGG TTACAGTTTG AGCAGGATAT TTGGTCCTGT AGTTTGCTAA CACACCCTGC 6060

AGCTCCAAAG GTTCCCCACC AACAGCAAAA AAATGAAAAT TTGACCCTTG AATGGGTTTT 6120

CCAGCACCAT TTTCATGAGT TTTTTGTGTC CCTGAATGCA AGTTTAACAT AGCAGTTACC 6180

CCAATAACCT CAGTTTTAAC AGTAACAGCT TCCCACATCA AAATATTTCC ACAGGTTAAG 6240

TCCTCATTTA AATTAGGCAA AGGAA 6265

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6254 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATACAC TCCGCTATCG CTACGTGACT GGGTCATGGC TGCGCCCCGA 2160

CACCCGCCAA CACCCGCTGA CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC 2220

AGACAAGCTG TGACCGTCTC CGGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG 2280

AAACGCGCGA GGCAGCGGAT CATAATCAGC CATACCACAT TTGTAGAGGT TTTACTTGCT 2340

TTAAAAAACC TCCCACACCT CCCCCTGAAC CTGAAACATA AAATGAATGC AATTGTTGTT 2400

GTTAACTTGT TTATTGCAGC TTATAATGGT TACAAATAAA GCAATAGCAT CACAAATTTC 2460

ACAAATAAAG CATTTTTTTC ACTGCATTCT AGTTGTGGTT TGTCCAAACT CATCAATGTA 2520

TCTTATCATG TCTGGATCAT AATCAGCCAT ACCACATTTG TAGAGGTTTT ACTTGCTTTA 2580

AAAAACCTCC CACACCTCCC CCTGAACCTG AAACATAAAA TGAATGCAAT TGTTGTTGTT 2640

AACTTGTTTA TTGCAGCTTA TAATGGTTAC AAATAAAGCA ATAGCATCAC AAATTTCACA 2700

AATAAAGCAT TTTTTTCACT GCATTCTAGT TGTGGTTTGT CCAAACTCAT CAATGTATCT 2760

TATCATGTCT GGATCCCAGC CAGACAAGGT TGTTGACACA AGACCCACAT CTGGTATAAA 2820

AGGAGGCAGT GGCCCACAGA GGAGCACAGC TGTGTTTGGC TGCAGGGCCA AGAGCGCTGT 2880

CAAGAAGACC CACACGCCCC CCTCCAGCAG CTGAATTCCA GCTGGCATTC CGGTACTGTT 2940

GGTAAAATGG AAGACGCCAA AAACATAAAG AAAGGCCCGG CGCCATTCTA TCCTCTAGAG 3000

GATGGAACCG CTGGAGAGCA ACTGCATAAG GCTATGAAGA GATACGCCCT GGTTCCTGGA 3060

ACAATTGCTT TTACAGATGC ACATATCGAG GTGAACATCA CGTACGCGGA ATACTTCGAA 3120

ATGTCCGTTC GGTTGGCAGA AGCTATGAAA CGATATGGGC TGAATACAAA TCACAGAATC 3180

GTCGTATGCA GTGAAAACTC TCTTCAATTC TTTATGCCGG TGTTGGGCGC GTTATTTATC 3240

GGAGTTGCAG TTGCGCCCGC GAACGACATT TATAATGAAC GTGAATTGCT CAACAGTATG 3300

AACATTTCGC AGCCTACCGT AGTGTTTGTT TCCAAAAAGG GGTTGCAAAA AATTTTGAAC 3360

GTGCAAAAAA AATTACCAAT AATCCAGAAA ATTATTATCA TGGATTCTAA AACGGATTAC 3420

CAGGGATTTC AGTCGATGTA CACGTTCGTC ACATCTCATC TACCTCCCGG TTTTAATGAA 3480

TACGATTTTG TACCAGAGTC CTTTGATCGT GACAAAACAA TTGCACTGAT AATGAATTCC 3540

TCTGGATCTA CTGGGTTACC TAAGGGTGTG GCCCTTCCGC ATAGAACTGC CTGCGTCAGA 3600

TTCTCGCATG CCAGAGATCC TATTTTTGGC AATCAAATCA TTCCGGATAC TGCGATTTTA 3660

AGTGTTGTTC CATTCCATCA CGGTTTTGGA ATGTTTACTA CACTCGGATA TTTGATATGT 3720

GGATTTCGAG TCGTCTTAAT GTATAGATTT GAAGAAGAGC TGTTTTTACG ATCCCTTCAG 3780

GATTACAAAA TTCAAAGTGC GTTGCTAGTA CCAACCCTAT TTTCATTCTT CGCCAAAAGC 3840

ACTCTGATTG ACAAATACGA TTTATCTAAT TTACACGAAA TTGCTTCTGG GGGCGCACCT 3900

CTTTCGAAAG AAGTCGGGGA AGCGGTTGCA AAACGCTTCC ATCTTCCAGG GATACGACAA 3960

GGATATGGGC TCACTGAGAC TACATCAGCT ATTCTGATTA CACCCGAGGG GGATGATAAA 4020

CCGGGCGCGG TCGGTAAAGT TGTTCCATTT TTTGAAGCGA AGGTTGTGGA TCTGGATACC 4080

GGGAAAACGC TGGGCGTTAA TCAGAGAGGC GAATTATGTG TCAGAGGACC TATGATTATG 4140

TCCGGTTATG TAAACAATCC GGAAGCGACC AACGCCTTGA TTGACAAGGA TGGATGGCTA 4200

CATTCTGGAG ACATAGCTTA CTGGGACGAA GACGAACACT TCTTCATAGT TGACCGCTTG 4260

AAGTCTTTAA TTAAATACAA AGGATATCAG GTGGCCCCCG CTGAATTGGA ATCGATATTG 4320

TTACAACACC CCAACATCTT CGACGCGGGC GTGGCAGGTC TTCCCGACGA TGACGCCGGT 4380

GAACTTCCCG CCGCCGTTGT TGTTTTGGAG CACGGAAAGA CGATGACGGA AAAAGAGATC 4440

GTGGATTACG TCGCCAGTCA AGTAACAACC GCGAAAAAGT TGCGCGGAGG AGTTGTGTTT 4500

GTGGACGAAG TACCGAAAGG TCTTACCGGA AAACTCGACG CAAGAAAAAT CAGAGAGATC 4560

CTCATAAAGG CCAAGAAGGG CGGAAAGTCC AAATTGTAAA ATGTAACTGT ATTCAGCGAT 4620

GACGAAATTC TTAGCTATTG TAATGACTCT AGAGGATCTT TGTGAAGGAA CCTTACTTCT 4680

GTGGTGTGAC ATAATTGGAC AAACTACCTA CAGAGATTTA AAGCTCTAAG GTAAATATAA 4740

AATTTTTAAG TGTATAATGT GTTAAACTAC TGATTCTAAT TGTTTGTGTA TTTTAGATTC 4800

CAACCTATGG AACTGATGAA TGGGAGCAGT GGTGGAATGC CTTTAATGAG GAAAACCTGT 4860

TTTGCTCAGA AGAAATGCCA TCTAGTGATG ATGAGGCTAC TGCTGACTCT CAACATTCTA 4920

CTCCTCCAAA AAAGAAGAGA AAGGTAGAAG ACCCCAAGGA CTTTCCTTCA GAATTGCTAA 4980

GTTTTTTGAG TCATGCTGTG TTTAGTAATA GAACTCTTGC TTGCTTTGCT ATTTACACCA 5040

CAAAGGAAAA AGCTGCACTG CTATACAAGA AAATTATGGA AAAATATTCT GTAACCTTTA 5100

TAAGTAGGCA TAACAGTTAT AATCATAACA TACTGTTTTT TCTTACTCCA CACAGGCATA 5160

GAGTGTCTGC TATTAATAAC TATGCTCAAA AATTGTGTAC CTTTAGCTTT TTAATTTGTA 5220

AAGGGGTTAA TAAGGAATAT TTGATGTATA GTGCCTTGAC TAGAGATCAT AATCAGCCAT 5280

ACCACATTTG TAGAGGTTTT ACTTGCTTTA AAAAACCTCC CACACCTCCC CCTGAACCTG 5340

AAACATAAAA TGAATGCAAT TGTTGTTGTT AACTTGTTTA TTGCAGCTTA TAATGGTTAC 5400

AAATAAAGCA ATAGCATCAC AAATTTCACA AATAAAGCAT TTTTTTCACT GCATTCTAGT 5460

TGTGGTTTGT CCAAACTCAT CAATGTATCT TATCATGTCT GGATCCCCAG GAAGCTCCTC 5520

TGTGTCCTCA TAAACCCTAA CCTCCTCTAC TTGAGAGGAC ATTCCAATCA TAGGCTGCCC 5580

ATCCACCCTC TGTGTCCTCC TGTTAATTAG GTCACTTAAC AAAAAGGAAA TTGGGTAGGG 5640

GTTTTTCACA GACCGCTTTC TAAGGGTAAT TTTAAAATAT CTGGGAAGTC CCTTCCACTG 5700

CTGTGTTCCA GAAGTGTTGG TAAACAGCCC ACAAATGTCA ACAGCAGAAA CATACAAGCT 5760

GTCAGCTTTG CACAAGGGCC CAACACCCTG CTCAGCAAGA AGCACTGTGG TTGCTGTGTT 5820

AGTAATGTGC AAAACAGGAG GCACATTTTC CCCACCTGTG TAGGTTCCAA AATATCTAGT 5880

GTTTTCATTT TTACTTGGAT CAGGAACCCA GCACTCCACT GGATAAGCAT TATCCTTATC 5940

CAAAACAGCC TTGTGGTCAG TGTTCATCTG CTGACTGTCA ACTGTAGCAT TTTTTGGGGT 6000

TACAGTTTGA GCAGGATATT TGGTCCTGTA GTTTGCTAAC ACACCCTGCA GCTCCAAAGG 6060

TTCCCCACCA ACAGCAAAAA AATGAAAATT TGACCCTTGA ATGGGTTTTC CAGCACCATT 6120

TTCATGAGTT TTTTGTGTCC CTGAATGCAA GTTTAACATA GCAGTTACCC CAATAACCTC 6180

AGTTTTAACA GTAACAGCTT CCCACATCAA AATATTTCCA CAGGTTAAGT CCTCATTTAA 6240

ATTAGGCAAA GGAA 6254

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1442 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGTACCCAGG CTGCATAACC AGGAGGTGAG TGGCAGGTGA GTGAAATTTC ATCTGTAGTT 60

ACAGCCACTC CTCATCACTC GCATTACCAC CAGAGCTCCA CTCCCTGTCA GATCAGCGGC 120

GGCATTAGAT TCTCATAGGA GCTCGAACCC TATTCTAAAC TGTTCATGTG AGGGATCTAG 180

GTTGCAAGCT CCCTATGAGA ATCTAATGCC TGATGATCTG TCACGGTCTC CCATCACCCC 240

TAGATGGGAC CATCTAGTTG CAGGAAAACA AGCTCAGGCT CCCACTGATT CTACACGATG 300

GTGAATTGTG GAATTATTTC ATTATATATA TTACAATGTA ATAATAATAG AAATAAAGCA 360

CACAATAAAT GTAATGTGCT TGAATCATCC CGAAACCATC CCACCCTGGT CTGTGAAAAA 420

ATTGTCTTCC ATGAAACCAG TCCCTGGTGC CAAAAACGTT GAGGACCACT GCTCCACAGA 480

ATCTATCGGT CACTCTTCCT CCCCTCACCC CCTTGCCCTA AAAGCACACC CTGCAAACCT 540

GCCATGAATT GACACTCTGT TTCTATCCCT TTTCCCCTTG TGTCTGTGTC TGGAGGAAGA 600

GGATAAAGGA CAAGCTGCCC CAAGTCCTAG CGGGCAGCTC GAGGAAGTGA AACTTACACG 660

TTGGTCTCCT GTTTCCTTAC CAAGCTTACC ATGGTAACCC CTGGTCCCGT TCAGCCACCA 720

CCACCCCACC CAGCACACCT CCAACCTCAG CCAGACAAGG TTGTTGACAC AAGAGAGCCC 780

TCAGGGGCAC AGAGAGAGTC TGGACACGTG GGGAGTCAGC CGTGTATCAT CGGAGGCGGC 840

CGGGCACATG GCAGGGATGA GGGAAAGACC AAGAGTCCTC TGTTGGGCCC AAGTCCTAGA 900

CAGACAAAAC CTAGACAATC ACGTGGCTGG CTGCATGCCT GTGGCTGTTG GGCTGGGCAG 960

GAGGAGGGAG GGGCGCTCTT TCCTGGAGGT GGTCCAGAGC ACCGGGTGGA CAGCCCTGGG 1020

GGAAAACTTC CACGTTTTGA TGGAGGTTAT CTTTGATAAC TCCACAGTGA CCTGGTTCGC 1080

CAAAGGAAAA GCAGGCAACG TGAGCTGTTT TTTTTTTCTC CAAGCTGAAC ACTAGGGGTC 1140

CTAGGCTTTT TGGGTCACCC GGCATGGCAG ACAGTCAACC TGGCAGGACA TCCGGGAGAG 1200

ACAGACACAG GCAGAGGGCA GAAAGGTCAA GGGAGGTTCT CAGGCCAAGG CTATTGGGGT 1260

TTGCTCAATT GTTCCTGAAT GCTCTTACAC ACGTACACAC ACAGAGCAGC ACACACACAC 1320

ACACACACAT GCCTCAGCAA GTCCCAGAGA GGGAGGTGTC GAGGGGGACC CGCTGGCTGT 1380

TCAGACGGAC TCCCAGAGCC AGTGAGTGGG TGGGGCTGGA ACATGAGTTC ATCTATTTCC 1440

TG 1442

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 761 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AAGCTTACCA TGGTAACCCC TGGTCCCGTT CAGCCACCAC CACCCCACCC AGCACACCTC 60

CAACCTCAGC CAGACAAGGT TGTTGACACA AGAGAGCCCT CAGGGGCACA GAGAGAGTCT 120

GGACACGTGG GGAGTCAGCC GTGTATCATC GGAGGCGGCC GGGCACATGG CAGGGATGAG 180

GGAAAGACCA AGAGTCCTCT GTTGGGCCCA AGTCCTAGAC AGACAAAACC TAGACAATCA 240

CGTGGCTGGC TGCATGCCTG TGGCTGTTGG GCTGGGCAGG AGGAGGGAGG GGCGCTCTTT 300

CCTGGAGGTG GTCCAGAGCA CCGGGTGGAC AGCCCTGGGG GAAAACTTCC ACGTTTTGAT 360

GGAGGTTATC TTTGATAACT CCACAGTGAC CTGGTTCGCC AAAGGAAAAG CAGGCAACGT 420

GAGCTGTTTT TTTTTTCTCC AAGCTGAACA CTAGGGGTCC TAGGCTTTTT GGGTCACCCG 480

GCATGGCAGA CAGTCAACCT GGCAGGACAT CCGGGAGAGA CAGACACAGG CAGAGGGCAG 540

AAAGGTCAAG GGAGGTTCTC AGGCCAAGGC TATTGGGGTT TGCTCAATTG TTCCTGAATG 600

CTCTTACACA CGTACACACA CAGAGCAGCA CACACACACA CACACACATG CCTCAGCAAG 660

TCCCAGAGAG GGAGGTGTCG AGGGGGACCC GCTGGCTGTT CAGACGGACT CCCAGAGCCA 720

GTGAGTGGGT GGGGCTGGAA CATGAGTTCA TCTATTTCCT G 761

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 165 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AAGCTTACCA TGGTAACCCC TGGTCCCGTT CAGCCACCAC CACCCCACCC AGCACACCTC 60

CAACCTCAGC CAGACAAGGT TGTTGACACA AGAGAGCCCT CAGGGGCACA GAGAGAGTCT 120

GGACACGTGG GGAGTCAGCC GTGTATCATC GGAGGCGGCC GGGCA 165

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGTTCATCTA TTTCCT 16

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTGGGGAGTC AGCCGTGTAT CATCG 25

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTCCAACCTC AGCCAGACAA GGTTGTTGAC ACAAGA 36

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCCAGACAAG GTTGTTGACA CAAGA 25

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 115 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCCACATCTG GTATAAAAGG AGGCAGTGGC CCACAGAGGA GCACAGCTGT GTTTGGCTGC 60

AGGGCCAAGA GCGCTGTCAA GAAGACCCAC ACGCCCCCCT CCAGCAGCTG AATTC 115

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 345 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGCCAGACGC CAACAAGGTA GGAGCTGGAG CATTCGGGCT GGGTTTCACC CCACCGCACG 60

GAGGCCTTTT GGGGTGGAGC CCTCAGGCTC AGGGCATACT ACAAACTTTG CCAGCAAATC 120

CGCCTCCTGC CTCCACCAAT CGCCAGTCAG GAAGGCAGCC TACCCCGCTG TCTCCACCTT 180

TGAGAAACAC TCATCCTCAG GCCATGCAGT GGAATTCCAC AACCTTCCAC CAAACTCTGC 240

AAGATCCCAG AGTGAGAGGC CTGTATTTCC CTGCTGGTGG CTCCAGTTCA GGAACAGTAA 300

ACCCTGTTCT GACTACTGCC TCTCCCTTAT CGTCAATCTT CTCGA 345

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4302 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TCGACCTCGA GGGATCTTTG TGAAGGAACC TTACTTCTGT GGTGTGACAT AATTGGACAA 60

ACTACCTACA GAGATTTAAA GCTCTAAGGT AAATATAAAA TTTTTAAGTG TATAATGTGT 120

TAAACTACTG ATTCTAATTG TTTGTGTATT TTAGATTCCA ACCTATGGAA CTGATGAATG 180

GGAGCAGTGG TGGAATGCCT TTAATGAGGA AAACCTGTTT TGCTCAGAAG AAATGCCATC 240

TAGTGATGAT GAGGCTACTG CTGACTCTCA ACATTCTACT CCTCCAAAAA AGAAGAGAAA 300

GGTAGAAGAC CCCAAGGACT TTCCTTCAGA ATTGCTAAGT TTTTTGAGTC ATGCTGTGTT 360

TAGTAATAGA ACTCTTGCTT GCTTTGCTAT TTACACCACA AAGGAAAAAG CTGCACTGCT 420

ATACAAGAAA ATTATGGAAA AATATTCTGT AACCTTTATA AGTAGGCATA ACAGTTATAA 480

TCATAACATA CTGTTTTTTC TTACTCCACA CAGGCATAGA GTGTCTGCTA TTAATAACTA 540

TGCTCAAAAA TTGTGTACCT TTAGCTTTTT AATTTGTAAA GGGGTTAATA AGGAATATTT 600

GATGTATAGT GCCTTGACTA GAGATCATAA TCAGCCATAC CACATTTGTA GAGGTTTTAC 660

TTGCTTTAAA AAACCTCCCA CACCTCCCCC TGAACCTGAA ACATAAAATG AATGCAATTG 720

TTGTTGTTAA CTTGTTTATT GCAGCTTATA ATGGTTACAA ATAAAGCAAT AGCATCACAA 780

ATTTCACAAA TAAAGCATTT TTTTCACTGC ATTCTAGTTG TGGTTTGTCC AAACTCATCA 840

ATGTATCTTA TCATGTCTGG ATCCGGCTGT GGAATGTGTG TCAGTTAGGG TGTGGAAAGT 900

CCCCAGGCTC CCCAGCAGGC AGAAGTATGC AAAGCATGCA TCTCAATTAG TCAGCAACCA 960

GGTGTGGAAA GTCCCCAGGC TCCCCAGCAG GCAGAAGTAT GCAAAGCATG CATCTCAATT 1020

AGTCAGCAAC CATAGTCCCG CCCCTAACTC CGCCCATCCC GCCCCTAACT CCGCCCAGTT 1080

CCGCCCATTC TCCGCCCCAT GGCTGACTAA TTTTTTTTAT TTATGCAGAG GCCGAGGCCG 1140

CCTCGGCCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT TTTTGGAGGC CTAGGCTTTT 1200

GCAAAAAGCT TCACGCTGCC GCAAGCACTC AGGGCGCAAG GGCTGCTAAA GGAAGCGGAA 1260

CACGTAGAAA GCCAGTCCGC AGAAACGGTG CTGACCCCGG ATGAATGTCA GCTACTGGGC 1320

TATCTGGACA AGGGAAAACG CAAGCGCAAA GAGAAAGCAG GTAGCTTGCA GTGGGCTTAC 1380

ATGGCGATAG CTAGACTGGG CGGTTTTATG GACAGCAAGC GAACCGGAAT TGCCAGCTGG 1440

GGCGCCCTCT GGTAAGGTTG GGAAGCCCTG CAAAGTAAAC TGGATGGCTT TCTTGCCGCC 1500

AAGGATCTGA TGGCGCAGGG GATCAAGATC TGATCAAGAG ACAGGATGAG GATCGTTTCG 1560

CATGATTGAA CAAGATGGAT TGCACGCAGG TTCTCCGGCC GCTTGGGTGG AGAGGCTATT 1620

CGGCTATGAC TGGGCACAAC AGACAATCGG CTGCTCTGAT GCCGCCGTGT TCCGGCTGTC 1680

AGCGCAGGGG CGCCCGGTTC TTTTTGTCAA GACCGACCTG TCCGGTGCCC TGAATGAACT 1740

GCAGGACGAG GCAGCGCGGC TATCGTGGCT GGCCACGACG GGCGTTCCTT GCGCAGCTGT 1800

GCTCGACGTT GTCACTGAAG CGGGAAGGGA CTGGCTGCTA TTGGGCGAAG TGCCGGGGCA 1860

GGATCTCCTG TCATCTCACC TTGCTCCTGC CGAGAAAGTA TCCATCATGG CTGATGCAAT 1920

GCGGCGGCTG CATACGCTTG ATCCGGCTAC CTGCCCATTC GACCACCAAG CGAAACATCG 1980

CATCGAGCGA GCACGTACTC GGATGGAAGC CGGTCTTGTC GATCAGGATG ATCTGGACGA 2040

AGAGCATCAG GGGCTCGCGC CAGCCGAACT GTTCGCCAGG CTCAAGGCGC GCATGCCCGA 2100

CGGCGAGGAT CTCGTCGTGA CCCATGGCGA TGCCTGCTTG CCGAATATCA TGGTGGAAAA 2160

TGGCCGCTTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT GTGGCGGACC GCTATCAGGA 2220

CATAGCGTTG GCTACCCGTG ATATTGCTGA AGAGCTTGGC GGCGAATGGG CTGACCGCTT 2280

CCTCGTGCTT TACGGTATCG CCGCTCCCGA TTCGCAGCGC ATCGCCTTCT ATCGCCTTCT 2340

TGACGAGTTC TTCTGAGCGG GACTCTGGGG TTCGAAATGA CCGACCAAGC GACGCCCAAC 2400

CTGCCATCAC GAGATTTCGA TTCCACCGCC GCCTTCTATG AAAGGTTGGG CTTCGGAATC 2460

GTTTTCCGGG ACGCCGGCTG GATGATCCTC CAGCGCGGGG ATCTCATGCT GGAGTTCTTC 2520

GCCCACCCCG GGCTCGATCC CCTCGCGAGT TGGTTCAGCT GCTGCCTGAG GCTGGACGAC 2580

CTCGCGGAGT TCTACCGGCA GTGCAAATCC GTCGGCATCC AGGAAACCAG CAGCGGCTAT 2640

CCGCGCATCC ATGCCCCCGA ACTGCAGGAG TGGGGAGGCA CGATGGCCGC TTTGGTCCCG 2700

GATCTTTGTG AAGGAACCTT ACTTCTGTGG TGTGACATAA TTGGACAAAC TACCTACAGA 2760

GATTTAAAGC TCTAAGGTAA ATATAAAATT TTTAAGTGTA TAATGTGTTA AACTACTGAT 2820

TCTAATTGTT TGTGTATTTT AGATTCCAAC CTATGGAACT GATGAATGGG AGCAGTGGTG 2880

GAATGCCTTT AATGAGGAAA ACCTGTTTTG CTCAGAAGAA ATGCCATCTA GTGATGATGA 2940

GGCTACTGCT GACTCTCAAC ATTCTACTCC TCCAAAAAAG AAGAGAAAGG TAGAAGACCC 3000

CAAGGACTTT CCTTCAGAAT TGCTAAGTTT TTTGAGTCAT GCTGTGTTTA GTAATAGAAC 3060

TCTTGCTTGC TTTGCTATTT ACACCACAAA GGAAAAAGCT GCACTGCTAT ACAAGAAAAT 3120

TATGGAAAAA TATTCTGTAA CCTTTATAAG TAGGCATAAC AGTTATAATC ATAACATACT 3180

GTTTTTTCTT ACTCCACACA GGCATAGAGT GTCTGCTATT AATAACTATG CTCAAAAATT 3240

GTGTACCTTT AGCTTTTTAA TTTGTAAAGG GGTTAATAAG GAATATTTGA TGTATAGTGC 3300

CTTGACTAGA GATCATAATC AGCCATACCA CATTTGTAGA GGTTTTACTT GCTTTAAAAA 3360

ACCTCCCACA CCTCCCCCTG AACCTGAAAC ATAAAATGAA TGCAATTGTT GTTGTTAACT 3420

TGTTTATTGC AGCTTATAAT GGTTACAAAT AAAGCAATAG CATCACAAAT TTCACAAATA 3480

AAGCATTTTT TTCACTGCAT TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTATC 3540

ATGTCTGGAT CCCCAGGAAG CTCCTCTGTG TCCTCATAAA CCCTAACCTC CTCTACTTGA 3600

GAGGACATTC CAATCATAGG CTGCCCATCC ACCCTCTGTG TCCTCCTGTT AATTAGGTCA 3660

CTTAACAAAA AGGAAATTGG GTAGGGGTTT TTCACAGACC GCTTTCTAAG GGTAATTTTA 3720

AAATATCTGG GAAGTCCCTT CCACTGCTGT GTTCCAGAAG TGTTGGTAAA CAGCCCACAA 3780

ATGTCAACAG CAGAAACATA CAAGCTGTCA GCTTTGCACA AGGGCCCAAC ACCCTGCTCA 3840

TCAAGAAGCA CTGTGGTTGC TGTGTTAGTA ATGTGCAAAA CAGGAGGCAC ATTTTCCCCA 3900

CCTGTGTAGG TTCCAAAATA TCTAGTGTTT TCATTTTTAC TTGGATCAGG AACCCAGCAC 3960

TCCACTGGAT AAGCATTATC CTTATCCAAA ACAGCCTTGT GGTCAGTGTT CATCTGCTGA 4020

CTGTCAACTG TAGCATTTTT TGGGGTTACA GTTTGAGCAG GATATTTGGT CCTGTAGTTT 4080

GCTAACACAC CCTGCAGCTC CAAAGGTTCC CCACCAACAG CAAAAAAATG AAAATTTGAC 4140

CCTTGAATGG GTTTTCCAGC ACCATTTTCA TGAGTTTTTT GTGTCCCTGA ATGCAAGTTT 4200

AACATAGCAG TTACCCCAAT AACCTCAGTT TTAACAGTAA CAGCTTCCCA CATCAAAATA 4260

TTTCCACAGG TTAAGTCCTC ATTTAAATTA GGCAAAGGAA TT 4302

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6170 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATACAC TCCGCTATCG CTACGTGACT GGGTCATGGC TGCGCCCCGA 2160

CACCCGCCAA CACCCGCTGA CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC 2220

AGACAAGCTG TGACCGTCTC CGGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG 2280

AAACGCGCGA GGCAGCGGAT CATAATCAGC CATACCACAT TTGTAGAGGT TTTACTTGCT 2340

TTAAAAAACC TCCCACACCT CCCCCTGAAC CTGAAACATA AAATGAATGC AATTGTTGTT 2400

GTTAACTTGT TTATTGCAGC TTATAATGGT TACAAATAAA GCAATAGCAT CACAAATTTC 2460

ACAAATAAAG CATTTTTTTC ACTGCATTCT AGTTGTGGTT TGTCCAAACT CATCAATGTA 2520

TCTTATCATG TCTGGATCAT AATCAGCCAT ACCACATTTG TAGAGGTTTT ACTTGCTTTA 2580

AAAAACCTCC CACACCTCCC CCTGAACCTG AAACATAAAA TGAATGCAAT TGTTGTTGTT 2640

AACTTGTTTA TTGCAGCTTA TAATGGTTAC AAATAAAGCA ATAGCATCAC AAATTTCACA 2700

AATAAAGCAT TTTTTTCACT GCATTCTAGT TGTGGTTTGT CCAAACTCAT CAATGTATCT 2760

TATCATGTCT GGATCCCAAG CTTGCATGCC TGCAGGTCGA CTCTAGAGGA TCCCCGGGTA 2820

CCGAGCTCGA ATTCCAGCTG GCATTCCGGT ACTGTTGGTA AAATGGAAGA CGCCAAAAAC 2880

ATAAAGAAAG GCCCGGCGCC ATTCTATCCT CTAGAGGATG GAACCGCTGG AGAGCAACTG 2940

CATAAGGCTA TGAAGAGATA CGCCCTGGTT CCTGGAACAA TTGCTTTTAC AGATGCACAT 3000

ATCGAGGTGA ACATCACGTA CGCGGAATAC TTCGAAATGT CCGTTCGGTT GGCAGAAGCT 3060

ATGAAACGAT ATGGGCTGAA TACAAATCAC AGAATCGTCG TATGCAGTGA AAACTCTCTT 3120

CAATTCTTTA TGCCGGTGTT GGGCGCGTTA TTTATCGGAG TTGCAGTTGC GCCCGCGAAC 3180

GACATTTATA ATGAACGTGA ATTGCTCAAC AGTATGAACA TTTCGCAGCC TACCGTAGTG 3240

TTTGTTTCCA AAAAGGGGTT GCAAAAAATT TTGAACGTGC AAAAAAAATT ACCAATAATC 3300

CAGAAAATTA TTATCATGGA TTCTAAAACG GATTACCAGG GATTTCAGTC GATGTACACG 3360

TTCGTCACAT CTCATCTACC TCCCGGTTTT AATGAATACG ATTTTGTACC AGAGTCCTTT 3420

GATCGTGACA AAACAATTGC ACTGATAATG AATTCCTCTG GATCTACTGG GTTACCTAAG 3480

GGTGTGGCCC TTCCGCATAG AACTGCCTGC GTCAGATTCT CGCATGCCAG AGATCCTATT 3540

TTTGGCAATC AAATCATTCC GGATACTGCG ATTTTAAGTG TTGTTCCATT CCATCACGGT 3600

TTTGGAATGT TTACTACACT CGGATATTTG ATATGTGGAT TTCGAGTCGT CTTAATGTAT 3660

AGATTTGAAG AAGAGCTGTT TTTACGATCC CTTCAGGATT ACAAAATTCA AAGTGCGTTG 3720

CTAGTACCAA CCCTATTTTC ATTCTTCGCC AAAAGCACTC TGATTGACAA ATACGATTTA 3780

TCTAATTTAC ACGAAATTGC TTCTGGGGGC GCACCTCTTT CGAAAGAAGT CGGGGAAGCG 3840

GTTGCAAAAC GCTTCCATCT TCCAGGGATA CGACAAGGAT ATGGGCTCAC TGAGACTACA 3900

TCAGCTATTC TGATTACACC CGAGGGGGAT GATAAACCGG GCGCGGTCGG TAAAGTTGTT 3960

CCATTTTTTG AAGCGAAGGT TGTGGATCTG GATACCGGGA AAACGCTGGG CGTTAATCAG 4020

AGAGGCGAAT TATGTGTCAG AGGACCTATG ATTATGTCCG GTTATGTAAA CAATCCGGAA 4080

GCGACCAACG CCTTGATTGA CAAGGATGGA TGGCTACATT CTGGAGACAT AGCTTACTGG 4140

GACGAAGACG AACACTTCTT CATAGTTGAC CGCTTGAAGT CTTTAATTAA ATACAAAGGA 4200

TATCAGGTGG CCCCCGCTGA ATTGGAATCG ATATTGTTAC AACACCCCAA CATCTTCGAC 4260

GCGGGCGTGG CAGGTCTTCC CGACGATGAC GCCGGTGAAC TTCCCGCCGC CGTTGTTGTT 4320

TTGGAGCACG GAAAGACGAT GACGGAAAAA GAGATCGTGG ATTACGTCGC CAGTCAAGTA 4380

ACAACCGCGA AAAAGTTGCG CGGAGGAGTT GTGTTTGTGG ACGAAGTACC GAAAGGTCTT 4440

ACCGGAAAAC TCGACGCAAG AAAAATCAGA GAGATCCTCA TAAAGGCCAA GAAGGGCGGA 4500

AAGTCCAAAT TGTAAAATGT AACTGTATTC AGCGATGACG AAATTCTTAG CTATTGTAAT 4560

GACTCTAGAG GATCTTTGTG AAGGAACCTT ACTTCTGTGG TGTGACATAA TTGGACAAAC 4620

TACCTACAGA GATTTAAAGC TCTAAGGTAA ATATAAAATT TTTAAGTGTA TAATGTGTTA 4680

AACTACTGAT TCTAATTGTT TGTGTATTTT AGATTCCAAC CTATGGAACT GATGAATGGG 4740

AGCAGTGGTG GAATGCCTTT AATGAGGAAA ACCTGTTTTG CTCAGAAGAA ATGCCATCTA 4800

GTGATGATGA GGCTACTGCT GACTCTCAAC ATTCTACTCC TCCAAAAAAG AAGAGAAAGG 4860

TAGAAGACCC CAAGGACTTT CCTTCAGAAT TGCTAAGTTT TTTGAGTCAT GCTGTGTTTA 4920

GTAATAGAAC TCTTGCTTGC TTTGCTATTT ACACCACAAA GGAAAAAGCT GCACTGCTAT 4980

ACAAGAAAAT TATGGAAAAA TATTCTGTAA CCTTTATAAG TAGGCATAAC AGTTATAATC 5040

ATAACATACT GTTTTTTCTT ACTCCACACA GGCATAGAGT GTCTGCTATT AATAACTATG 5100

CTCAAAAATT GTGTACCTTT AGCTTTTTAA TTTGTAAAGG GGTTAATAAG GAATATTTGA 5160

TGTATAGTGC CTTGACTAGA GATCATAATC AGCCATACCA CATTTGTAGA GGTTTTACTT 5220

GCTTTAAAAA ACCTCCCACA CCTCCCCCTG AACCTGAAAC ATAAAATGAA TGCAATTGTT 5280

GTTGTTAACT TGTTTATTGC AGCTTATAAT GGTTACAAAT AAAGCAATAG CATCACAAAT 5340

TTCACAAATA AAGCATTTTT TTCACTGCAT TCTAGTTGTG GTTTGTCCAA ACTCATCAAT 5400

GTATCTTATC ATGTCTGGAT CCCCAGGAAG CTCCTCTGTG TCCTCATAAA CCCTAACCTC 5460

CTCTACTTGA GAGGACATTC CAATCATAGG CTGCCCATCC ACCCTCTGTG TCCTCCTGTT 5520

AATTAGGTCA CTTAACAAAA AGGAAATTGG GTAGGGGTTT TTCACAGACC GCTTTCTAAG 5580

GGTAATTTTA AAATATCTGG GAAGTCCCTT CCACTGCTGT GTTCCAGAAG TGTTGGTAAA 5640

CAGCCCACAA ATGTCAACAG CAGAAACATA CAAGCTGTCA GCTTTGCACA AGGGCCCAAC 5700

ACCCTGCTCA GCAAGAAGCA CTGTGGTTGC TGTGTTAGTA ATGTGCAAAA CAGGAGGCAC 5760

ATTTTCCCCA CCTGTGTAGG TTCCAAAATA TCTAGTGTTT TCATTTTTAC TTGGATCAGG 5820

AACCCAGCAC TCCACTGGAT AAGCATTATC CTTATCCAAA ACAGCCTTGT GGTCAGTGTT 5880

CATCTGCTGA CTGTCAACTG TAGCATTTTT TGGGGTTACA GTTTGAGCAG GATATTTGGT 5940

CCTGTAGTTT GCTAACACAC CCTGCAGCTC CAAAGGTTCC CCACCAACAG CAAAAAAATG 6000

AAAATTTGAC CCTTGAATGG GTTTTCCAGC ACCATTTTCA TGAGTTTTTT GTGTCCCTGA 6060

ATGCAAGTTT AACATAGCAG TTACCCCAAT AACCTCAGTT TTAACAGTAA CAGCTTCCCA 6120

CATCAAAATA TTTCCACAGG TTAAGTCCTC ATTTAAATTA GGCAAAGGAA 6170

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10533 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATTCGA CCTCGAGGGA TCTTTGTGAA GGAACCTTAC TTCTGTGGTG 2160

TGACATAATT GGACAAACTA CCTACAGAGA TTTAAAGCTC TAAGGTAAAT ATAAAATTTT 2220

TAAGTGTATA ATGTGTTAAA CTACTGATTC TAATTGTTTG TGTATTTTAG ATTCCAACCT 2280

ATGGAACTGA TGAATGGGAG CAGTGGTGGA ATGCCTTTAA TGAGGAAAAC CTGTTTTGCT 2340

CAGAAGAAAT GCCATCTAGT GATGATGAGG CTACTGCTGA CTCTCAACAT TCTACTCCTC 2400

CAAAAAAGAA GAGAAAGGTA GAAGACCCCA AGGACTTTCC TTCAGAATTG CTAAGTTTTT 2460

TGAGTCATGC TGTGTTTAGT AATAGAACTC TTGCTTGCTT TGCTATTTAC ACCACAAAGG 2520

AAAAAGCTGC ACTGCTATAC AAGAAAATTA TGGAAAAATA TTCTGTAACC TTTATAAGTA 2580

GGCATAACAG TTATAATCAT AACATACTGT TTTTTCTTAC TCCACACAGG CATAGAGTGT 2640

CTGCTATTAA TAACTATGCT CAAAAATTGT GTACCTTTAG CTTTTTAATT TGTAAAGGGG 2700

TTAATAAGGA ATATTTGATG TATAGTGCCT TGACTAGAGA TCATAATCAG CCATACCACA 2760

TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC TCCCCCTGAA CCTGAAACAT 2820

AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG CTTATAATGG TTACAAATAA 2880

AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT 2940

TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC GGCTGTGGAA TGTGTGTCAG 3000

TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC 3060

AATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC CAGCAGGCAG AAGTATGCAA 3120

AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC 3180

CTAACTCCGC CCAGTTCCGC CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT 3240

GCAGAGGCCG AGGCCGCCTC GGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT 3300

GGAGGCCTAG GCTTTTGCAA AAAGCTTCAC GCTGCCGCAA GCACTCAGGG CGCAAGGGCT 3360

GCTAAAGGAA GCGGAACACG TAGAAAGCCA GTCCGCAGAA ACGGTGCTGA CCCCGGATGA 3420

ATGTCAGCTA CTGGGCTATC TGGACAAGGG AAAACGCAAG CGCAAAGAGA AAGCAGGTAG 3480

CTTGCAGTGG GCTTACATGG CGATAGCTAG ACTGGGCGGT TTTATGGACA GCAAGCGAAC 3540

CGGAATTGCC AGCTGGGGCG CCCTCTGGTA AGGTTGGGAA GCCCTGCAAA GTAAACTGGA 3600

TGGCTTTCTT GCCGCCAAGG ATCTGATGGC GCAGGGGATC AAGATCTGAT CAAGAGACAG 3660

GATGAGGATC GTTTCGCATG ATTGAACAAG ATGGATTGCA CGCAGGTTCT CCGGCCGCTT 3720

GGGTGGAGAG GCTATTCGGC TATGACTGGG CACAACAGAC AATCGGCTGC TCTGATGCCG 3780

CCGTGTTCCG GCTGTCAGCG CAGGGGCGCC CGGTTCTTTT TGTCAAGACC GACCTGTCCG 3840

GTGCCCTGAA TGAACTGCAG GACGAGGCAG CGCGGCTATC GTGGCTGGCC ACGACGGGCG 3900

TTCCTTGCGC AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG 3960

GCGAAGTGCC GGGGCAGGAT CTCCTGTCAT CTCACCTTGC TCCTGCCGAG AAAGTATCCA 4020

TCATGGCTGA TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC 4080

ACCAAGCGAA ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC 4140

AGGATGATCT GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTC GCCAGGCTCA 4200

AGGCGCGCAT GCCCGACGGC GAGGATCTCG TCGTGACCCA TGGCGATGCC TGCTTGCCGA 4260

ATATCATGGT GGAAAATGGC CGCTTTTCTG GATTCATCGA CTGTGGCCGG CTGGGTGTGG 4320

CGGACCGCTA TCAGGACATA GCGTTGGCTA CCCGTGATAT TGCTGAAGAG CTTGGCGGCG 4380

AATGGGCTGA CCGCTTCCTC GTGCTTTACG GTATCGCCGC TCCCGATTCG CAGCGCATCG 4440

CCTTCTATCG CCTTCTTGAC GAGTTCTTCT GAGCGGGACT CTGGGGTTCG AAATGACCGA 4500

CCAAGCGACG CCCAACCTGC CATCACGAGA TTTCGATTCC ACCGCCGCCT TCTATGAAAG 4560

GTTGGGCTTC GGAATCGTTT TCCGGGACGC CGGCTGGATG ATCCTCCAGC GCGGGGATCT 4620

CATGCTGGAG TTCTTCGCCC ACCCCGGGCT CGATCCCCTC GCGAGTTGGT TCAGCTGCTG 4680

CCTGAGGCTG GACGACCTCG CGGAGTTCTA CCGGCAGTGC AAATCCGTCG GCATCCAGGA 4740

AACCAGCAGC GGCTATCCGC GCATCCATGC CCCCGAACTG CAGGAGTGGG GAGGCACGAT 4800

GGCCGCTTTG GTCCCGGATC TTTGTGAAGG AACCTTACTT CTGTGGTGTG ACATAATTGG 4860

ACAAACTACC TACAGAGATT TAAAGCTCTA AGGTAAATAT AAAATTTTTA AGTGTATAAT 4920

GTGTTAAACT ACTGATTCTA ATTGTTTGTG TATTTTAGAT TCCAACCTAT GGAACTGATG 4980

AATGGGAGCA GTGGTGGAAT GCCTTTAATG AGGAAAACCT GTTTTGCTCA GAAGAAATGC 5040

CATCTAGTGA TGATGAGGCT ACTGCTGACT CTCAACATTC TACTCCTCCA AAAAAGAAGA 5100

GAAAGGTAGA AGACCCCAAG GACTTTCCTT CAGAATTGCT AAGTTTTTTG AGTCATGCTG 5160

TGTTTAGTAA TAGAACTCTT GCTTGCTTTG CTATTTACAC CACAAAGGAA AAAGCTGCAC 5220

TGCTATACAA GAAAATTATG GAAAAATATT CTGTAACCTT TATAAGTAGG CATAACAGTT 5280

ATAATCATAA CATACTGTTT TTTCTTACTC CACACAGGCA TAGAGTGTCT GCTATTAATA 5340

ACTATGCTCA AAAATTGTGT ACCTTTAGCT TTTTAATTTG TAAAGGGGTT AATAAGGAAT 5400

ATTTGATGTA TAGTGCCTTG ACTAGAGATC ATAATCAGCC ATACCACATT TGTAGAGGTT 5460

TTACTTGCTT TAAAAAACCT CCCACACCTC CCCCTGAACC TGAAACATAA AATGAATGCA 5520

ATTGTTGTTG TTAACTTGTT TATTGCAGCT TATAATGGTT ACAAATAAAG CAATAGCATC 5580

ACAAATTTCA CAAATAAAGC ATTTTTTTCA CTGCATTCTA GTTGTGGTTT GTCCAAACTC 5640

ATCAATGTAT CTTATCATGT CTGGATCCCC AGGAAGCTCC TCTGTGTCCT CATAAACCCT 5700

AACCTCCTCT ACTTGAGAGG ACATTCCAAT CATAGGCTGC CCATCCACCC TCTGTGTCCT 5760

CCTGTTAATT AGGTCACTTA ACAAAAAAGGA AATTGGGTAG GGGTTTTTCA CAGACCGCTT 5820

TCTAAGGGTA ATTTTAAAAT ATCTGGGAAG TCCCTTCCAC TGCTGTGTTC CAGAAGTGTT 5880

GGTAAACAGC CCACAAATGT CAACAGCAGA AACATACAAG CTGTCAGCTT TGCACAAGGG 5940

CCCAACACCC TGCTCATCAA GAAGCACTGT GGTTGCTGTG TTAGTAATGT GCAAAACAGG 6000

AGGCACATTT TCCCCACCTG TGTAGGTTCC AAAATATCTA GTGTTTTCAT TTTTACTTGG 6060

ATCAGGAACC CAGCACTCCA CTGGATAAGC ATTATCCTTA TCCAAAACAG CCTTGTGGTC 6120

AGTGTTCATC TGCTGACTGT CAACTGTAGC ATTTTTTGGG GTTACAGTTT GAGCAGGATA 6180

TTTGGTCCTG TAGTTTGCTA ACACACCCTG CAGCTCCAAA GGTTCCCCAC CAACAGCAAA 6240

AAAATGAAAA TTTGACCCTT GAATGGGTTT TCCAGCACCA TTTTCATGAG TTTTTTGTGT 6300

CCCTGAATGC AAGTTTAACA TAGCAGTTAC CCCAATAACC TCAGTTTTAA CAGTAACAGC 6360

TTCCCACATC AAAATATTTC CACAGGTTAA GTCCTCATTT AAATTAGGCA AAGGAATTAT 6420

ACACTCCGCT ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG 6480

CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG 6540

TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGGCAGC 6600

GGATCATAAT CAGCCATACC ACATTTGTAG AGGTTTTACT TGCTTTAAAA AACCTCCCAC 6660

ACCTCCCCCT GAACCTGAAA CATAAAATGA ATGCAATTGT TGTTGTTAAC TTGTTTATTG 6720

CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT 6780

TTTCACTGCA TTCTAGTTGT GGTTTGTCCA AACTCATCAA TGTATCTTAT CATGTCTGGA 6840

TCATAATCAG CCATACCACA TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC 6900

TCCCCCTGAA CCTGAAACAT AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG 6960

CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT 7020

CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC 7080

CACCCACATC TGGTATAAAA GGAGGCAGTG GCCCACAGAG GAGCACAGCT GTGTTTGGCT 7140

GCAGGGCCAA GAGCGCTGTC AAGAAGACCC ACACGCCCCC CTCCAGCAGC TGAATTCCAG 7200

CTGGCATTCC GGTACTGTTG GTAAAATGGA AGACGCCAAA AACATAAAGA AAGGCCCGGC 7260

GCCATTCTAT CCTCTAGAGG ATGGAACCGC TGGAGAGCAA CTGCATAAGG CTATGAAGAG 7320

ATACGCCCTG GTTCCTGGAA CAATTGCTTT TACAGATGCA CATATCGAGG TGAACATCAC 7380

GTACGCGGAA TACTTCGAAA TGTCCGTTCG GTTGGCAGAA GCTATGAAAC GATATGGGCT 7440

GAATACAAAT CACAGAATCG TCGTATGCAG TGAAAACTCT CTTCAATTCT TTATGCCGGT 7500

GTTGGGCGCG TTATTTATCG GAGTTGCAGT TGCGCCCGCG AACGACATTT ATAATGAACG 7560

TGAATTGCTC AACAGTATGA ACATTTCGCA GCCTACCGTA GTGTTTGTTT CCAAAAAGGG 7620

GTTGCAAAAA ATTTTGAACG TGCAAAAAAA ATTACCAATA ATCCAGAAAA TTATTATCAT 7680

GGATTCTAAA ACGGATTACC AGGGATTTCA GTCGATGTAC ACGTTCGTCA CATCTCATCT 7740

ACCTCCCGGT TTTAATGAAT ACGATTTTGT ACCAGAGTCC TTTGATCGTG ACAAAACAAT 7800

TGCACTGATA ATGAATTCCT CTGGATCTAC TGGGTTACCT AAGGGTGTGG CCCTTCCGCA 7860

TAGAACTGCC TGCGTCAGAT TCTCGCATGC CAGAGATCCT ATTTTTGGCA ATCAAATCAT 7920

TCCGGATACT GCGATTTTAA GTGTTGTTCC ATTCCATCAC GGTTTTGGAA TGTTTACTAC 7980

ACTCGGATAT TTGATATGTG GATTTCGAGT CGTCTTAATG TATAGATTTG AAGAAGAGCT 8040

GTTTTTACGA TCCCTTCAGG ATTACAAAAT TCAAAGTGCG TTGCTAGTAC CAACCCTATT 8100

TTCATTCTTC GCCAAAAGCA CTCTGATTGA CAAATACGAT TTATCTAATT TACACGAAAT 8160

TGCTTCTGGG GGCGCACCTC TTTCGAAAGA AGTCGGGGAA GCGGTTGCAA AACGCTTCCA 8220

TCTTCCAGGG ATACGACAAG GATATGGGCT CACTGAGACT ACATCAGCTA TTCTGATTAC 8280

ACCCGAGGGG GATGATAAAC CGGGCGCGGT CGGTAAAGTT GTTCCATTTT TTGAAGCGAA 8340

GGTTGTGGAT CTGGATACCG GGAAAACGCT GGGCGTTAAT CAGAGAGGCG AATTATGTGT 8400

CAGAGGACCT ATGATTATGT CCGGTTATGT AAACAATCCG GAAGCGACCA ACGCCTTGAT 8460

TGACAAGGAT GGATGGCTAC ATTCTGGAGA CATAGCTTAC TGGGACGAAG ACGAACACTT 8520

CTTCATAGTT GACCGCTTGA AGTCTTTAAT TAAATACAAA GGATATCAGG TGGCCCCCGC 8580

TGAATTGGAA TCGATATTGT TACAACACCC CAACATCTTC GACGCGGGCG TGGCAGGTCT 8640

TCCCGACGAT GACGCCGGTG AACTTCCCGC CGCCGTTGTT GTTTTGGAGC ACGGAAAGAC 8700

GATGACGGAA AAAGAGATCG TGGATTACGT CGCCAGTCAA GTAACAACCG CGAAAAAGTT 8760

GCGCGGAGGA GTTGTGTTTG TGGACGAAGT ACCGAAAGGT CTTACCGGAA AACTCGACGC 8820

AAGAAAAATC AGAGAGATCC TCATAAAGGC CAAGAAGGGC GGAAAGTCCA AATTGTAAAA 8880

TGTAACTGTA TTCAGCGATG ACGAAATTCT TAGCTATTGT AATGACTCTA GAGGATCTTT 8940

GTGAAGGAAC CTTACTTCTG TGGTGTGACA TAATTGGACA AACTACCTAC AGAGATTTAA 9000

AGCTCTAAGG TAAATATAAA ATTTTTAAGT GTATAATGTG TTAAACTACT GATTCTAATT 9060

GTTTGTGTAT TTTAGATTCC AACCTATGGA ACTGATGAAT GGGAGCAGTG GTGGAATGCC 9120

TTTAATGAGG AAAACCTGTT TTGCTCAGAA GAAATGCCAT CTAGTGATGA TGAGGCTACT 9180

GCTGACTCTC AACATTCTAC TCCTCCAAAA AAGAAGAGAA AGGTAGAAGA CCCCAAGGAC 9240

TTTCCTTCAG AATTGCTAAG TTTTTTGAGT CATGCTGTGT TTAGTAATAG AACTCTTGCT 9300

TGCTTTGCTA TTTACACCAC AAAGGAAAAA GCTGCACTGC TATACAAGAA AATTATGGAA 9360

AAATATTCTG TAACCTTTAT AAGTAGGCAT AACAGTTATA ATCATAACAT ACTGTTTTTT 9420

CTTACTCCAC ACAGGCATAG AGTGTCTGCT ATTAATAACT ATGCTCAAAA ATTGTGTACC 9480

TTTAGCTTTT TAATTTGTAA AGGGGTTAAT AAGGAATATT TGATGTATAG TGCCTTGACT 9540

AGAGATCATA ATCAGCCATA CCACATTTGT AGAGGTTTTA CTTGCTTTAA AAAACCTCCC 9600

ACACCTCCCC CTGAACCTGA AACATAAAAT GAATGCAATT GTTGTTGTTA ACTTGTTTAT 9660

TGCAGCTTAT AATGGTTACA AATAAAGCAA TAGCATCACA AATTTCACAA ATAAAGCATT 9720

TTTTTCACTG CATTCTAGTT GTGGTTTGTC CAAACTCATC AATGTATCTT ATCATGTCTG 9780

GATCCCCAGG AAGCTCCTCT GTGTCCTCAT AAACCCTAAC CTCCTCTACT TGAGAGGACA 9840

TTCCAATCAT AGGCTGCCCA TCCACCCTCT GTGTCCTCCT GTTAATTAGG TCACTTAACA 9900

AAAAGGAAAT TGGGTAGGGG TTTTTCACAG ACCGCTTTCT AAGGGTAATT TTAAAATATC 9960

TGGGAAGTCC CTTCCACTGC TGTGTTCCAG AAGTGTTGGT AAACAGCCCA CAAATGTCAA 10020

CAGCAGAAAC ATACAAGCTG TCAGCTTTGC ACAAGGGCCC AACACCCTGC TCAGCAAGAA 10080

GCACTGTGGT TGCTGTGTTA GTAATGTGCA AAACAGGAGG CACATTTTCC CCACCTGTGT 10140

AGGTTCCAAA ATATCTAGTG TTTTCATTTT TACTTGGATC AGGAACCCAG CACTCCACTG 10200

GATAAGCATT ATCCTTATCC AAAACAGCCT TGTGGTCAGT GTTCATCTGC TGACTGTCAA 10260

CTGTAGCATT TTTTGGGGTT ACAGTTTGAG CAGGATATTT GGTCCTGTAG TTTGCTAACA 10320

CACCCTGCAG CTCCAAAGGT TCCCCACCAA CAGCAAAAAA ATGAAAATTT GACCCTTGAA 10380

TGGGTTTTCC AGCACCATTT TCATGAGTTT TTTGTGTCCC TGAATGCAAG TTTAACATAG 10440

CAGTTACCCC AATAACCTCA GTTTTAACAG TAACAGCTTC CCACATCAAA ATATTTCCAC 10500

AGGTTAAGTC CTCATTTAAA TTAGGCAAAG GAA 10533

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6229 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATACAC TCCGCTATCG CTACGTGACT GGGTCATGGC TGCGCCCCGA 2160

CACCCGCCAA CACCCGCTGA CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC 2220

AGACAAGCTG TGACCGTCTC CGGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG 2280

AAACGCGCGA GGCAGCGGAT CATAATCAGC CATACCACAT TTGTAGAGGT TTTACTTGCT 2340

TTAAAAAACC TCCCACACCT CCCCCTGAAC CTGAAACATA AAATGAATGC AATTGTTGTT 2400

GTTAACTTGT TTATTGCAGC TTATAATGGT TACAAATAAA GCAATAGCAT CACAAATTTC 2460

ACAAATAAAG CATTTTTTTC ACTGCATTCT AGTTGTGGTT TGTCCAAACT CATCAATGTA 2520

TCTTATCATG TCTGGATCAT AATCAGCCAT ACCACATTTG TAGAGGTTTT ACTTGCTTTA 2580

AAAAACCTCC CACACCTCCC CCTGAACCTG AAACATAAAA TGAATGCAAT TGTTGTTGTT 2640

AACTTGTTTA TTGCAGCTTA TAATGGTTAC AAATAAAGCA ATAGCATCAC AAATTTCACA 2700

AATAAAGCAT TTTTTTCACT GCATTCTAGT TGTGGTTTGT CCAAACTCAT CAATGTATCT 2760

TATCATGTCT GGATCCCACC CACATCTGGT ATAAAAGGAG GCAGTGGCCC ACAGAGGAGC 2820

ACAGCTGTGT TTGGCTGCAG GGCCAAGAGC GCTGTCAAGA AGACCCACAC GCCCCCCTCC 2880

AGCAGCTGAA TTCCAGCTGG CATTCCGGTA CTGTTGGTAA AATGGAAGAC GCCAAAAACA 2940

TAAAGAAAGG CCCGGCGCCA TTCTATCCTC TAGAGGATGG AACCGCTGGA GAGCAACTGC 3000

ATAAGGCTAT GAAGAGATAC GCCCTGGTTC CTGGAACAAT TGCTTTTACA GATGCACATA 3060

TCGAGGTGAA CATCACGTAC GCGGAATACT TCGAAATGTC CGTTCGGTTG GCAGAAGCTA 3120

TGAAACGATA TGGGCTGAAT ACAAATCACA GAATCGTCGT ATGCAGTGAA AACTCTCTTC 3180

AATTCTTTAT GCCGGTGTTG GGCGCGTTAT TTATCGGAGT TGCAGTTGCG CCCGCGAACG 3240

ACATTTATAA TGAACGTGAA TTGCTCAACA GTATGAACAT TTCGCAGCCT ACCGTAGTGT 3300

TTGTTTCCAA AAAGGGGTTG CAAAAAATTT TGAACGTGCA AAAAAAATTA CCAATAATCC 3360

AGAAAATTAT TATCATGGAT TCTAAAACGG ATTACCAGGG ATTTCAGTCG ATGTACACGT 3420

TCGTCACATC TCATCTACCT CCCGGTTTTA ATGAATACGA TTTTGTACCA GAGTCCTTTG 3480

ATCGTGACAA AACAATTGCA CTGATAATGA ATTCCTCTGG ATCTACTGGG TTACCTAAGG 3540

GTGTGGCCCT TCCGCATAGA ACTGCCTGCG TCAGATTCTC GCATGCCAGA GATCCTATTT 3600

TTGGCAATCA AATCATTCCG GATACTGCGA TTTTAAGTGT TGTTCCATTC CATCACGGTT 3660

TTGGAATGTT TACTACACTC GGATATTTGA TATGTGGATT TCGAGTCGTC TTAATGTATA 3720

GATTTGAAGA AGAGCTGTTT TTACGATCCC TTCAGGATTA CAAAATTCAA AGTGCGTTGC 3780

TAGTACCAAC CCTATTTTCA TTCTTCGCCA AAAGCACTCT GATTGACAAA TACGATTTAT 3840

CTAATTTACA CGAAATTGCT TCTGGGGGCG CACCTCTTTC GAAAGAAGTC GGGGAAGCGG 3900

TTGCAAAACG CTTCCATCTT CCAGGGATAC GACAAGGATA TGGGCTCACT GAGACTACAT 3960

CAGCTATTCT GATTACACCC GAGGGGGATG ATAAACCGGG CGCGGTCGGT AAAGTTGTTC 4020

CATTTTTTGA AGCGAAGGTT GTGGATCTGG ATACCGGGAA AACGCTGGGC GTTAATCAGA 4080

GAGGCGAATT ATGTGTCAGA GGACCTATGA TTATGTCCGG TTATGTAAAC AATCCGGAAG 4140

CGACCAACGC CTTGATTGAC AAGGATGGAT GGCTACATTC TGGAGACATA GCTTACTGGG 4200

ACGAAGACGA ACACTTCTTC ATAGTTGACC GCTTGAAGTC TTTAATTAAA TACAAAGGAT 4260

ATCAGGTGGC CCCCGCTGAA TTGGAATCGA TATTGTTACA ACACCCCAAC ATCTTCGACG 4320

CGGGCGTGGC AGGTCTTCCC GACGATGACG CCGGTGAACT TCCCGCCGCC GTTGTTGTTT 4380

TGGAGCACGG AAAGACGATG ACGGAAAAAG AGATCGTGGA TTACGTCGCC AGTCAAGTAA 4440

CAACCGCGAA AAAGTTGCGC GGAGGAGTTG TGTTTGTGGA CGAAGTACCG AAAGGTCTTA 4500

CCGGAAAACT CGACGCAAGA AAAATCAGAG AGATCCTCAT AAAGGCCAAG AAGGGCGGAA 4560

AGTCCAAATT GTAAAATGTA ACTGTATTCA GCGATGACGA AATTCTTAGC TATTGTAATG 4620

ACTCTAGAGG ATCTTTGTGA AGGAACCTTA CTTCTGTGGT GTGACATAAT TGGACAAACT 4680

ACCTACAGAG ATTTAAAGCT CTAAGGTAAA TATAAAATTT TTAAGTGTAT AATGTGTTAA 4740

ACTACTGATT CTAATTGTTT GTGTATTTTA GATTCCAACC TATGGAACTG ATGAATGGGA 4800

GCAGTGGTGG AATGCCTTTA ATGAGGAAAA CCTGTTTTGC TCAGAAGAAA TGCCATCTAG 4860

TGATGATGAG GCTACTGCTG ACTCTCAACA TTCTACTCCT CCAAAAAAGA AGAGAAAGGT 4920

AGAAGACCCC AAGGACTTTC CTTCAGAATT GCTAAGTTTT TTGAGTCATG CTGTGTTTAG 4980

TAATAGAACT CTTGCTTGCT TTGCTATTTA CACCACAAAG GAAAAAGCTG CACTGCTATA 5040

CAAGAAAATT ATGGAAAAAT ATTCTGTAAC CTTTATAAGT AGGCATAACA GTTATAATCA 5100

TAACATACTG TTTTTTCTTA CTCCACACAG GCATAGAGTG TCTGCTATTA ATAACTATGC 5160

TCAAAAATTG TGTACCTTTA GCTTTTTAAT TTGTAAAGGG GTTAATAAGG AATATTTGAT 5220

GTATAGTGCC TTGACTAGAG ATCATAATCA GCCATACCAC ATTTGTAGAG GTTTTACTTG 5280

CTTTAAAAAA CCTCCCACAC CTCCCCCTGA ACCTGAAACA TAAAATGAAT GCAATTGTTG 5340

TTGTTAACTT GTTTATTGCA GCTTATAATG GTTACAAATA AAGCAATAGC ATCACAAATT 5400

TCACAAATAA AGCATTTTTT TCACTGCATT CTAGTTGTGG TTTGTCCAAA CTCATCAATG 5460

TATCTTATCA TGTCTGGATC CCCAGGAAGC TCCTCTGTGT CCTCATAAAC CCTAACCTCC 5520

TCTACTTGAG AGGACATTCC AATCATAGGC TGCCCATCCA CCCTCTGTGT CCTCCTGTTA 5580

ATTAGGTCAC TTAACAAAAA GGAAATTGGG TAGGGGTTTT TCACAGACCG CTTTCTAAGG 5640

GTAATTTTAA AATATCTGGG AAGTCCCTTC CACTGCTGTG TTCCAGAAGT GTTGGTAAAC 5700

AGCCCACAAA TGTCAACAGC AGAAACATAC AAGCTGTCAG CTTTGCACAA GGGCCCAACA 5760

CCCTGCTCAG CAAGAAGCAC TGTGGTTGCT GTGTTAGTAA TGTGCAAAAC AGGAGGCACA 5820

TTTTCCCCAC CTGTGTAGGT TCCAAAATAT CTAGTGTTTT CATTTTTACT TGGATCAGGA 5880

ACCCAGCACT CCACTGGATA AGCATTATCC TTATCCAAAA CAGCCTTGTG GTCAGTGTTC 5940

ATCTGCTGAC TGTCAACTGT AGCATTTTTT GGGGTTACAG TTTGAGCAGG ATATTTGGTC 6000

CTGTAGTTTG CTAACACACC CTGCAGCTCC AAAGGTTCCC CACCAACAGC AAAAAAATGA 6060

AAATTTGACC CTTGAATGGG TTTTCCAGCA CCATTTTCAT GAGTTTTTTG TGTCCCTGAA 6120

TGCAAGTTTA ACATAGCAGT TACCCCAATA ACCTCAGTTT TAACAGTAAC AGCTTCCCAC 6180

ATCAAAATAT TTCCACAGGT TAAGTCCTCA TTTAAATTAG GCAAAGGAA 6229

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10768 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATTCGA CCTCGAGGGA TCTTTGTGAA GGAACCTTAC TTCTGTGGTG 2160

TGACATAATT GGACAAACTA CCTACAGAGA TTTAAAGCTC TAAGGTAAAT ATAAAATTTT 2220

TAAGTGTATA ATGTGTTAAA CTACTGATTC TAATTGTTTG TGTATTTTAG ATTCCAACCT 2280

ATGGAACTGA TGAATGGGAG CAGTGGTGGA ATGCCTTTAA TGAGGAAAAC CTGTTTTGCT 2340

CAGAAGAAAT GCCATCTAGT GATGATGAGG CTACTGCTGA CTCTCAACAT TCTACTCCTC 2400

CAAAAAAGAA GAGAAAGGTA GAAGACCCCA AGGACTTTCC TTCAGAATTG CTAAGTTTTT 2460

TGAGTCATGC TGTGTTTAGT AATAGAACTC TTGCTTGCTT TGCTATTTAC ACCACAAAGG 2520

AAAAAGCTGC ACTGCTATAC AAGAAAATTA TGGAAAAATA TTCTGTAACC TTTATAAGTA 2580

GGCATAACAG TTATAATCAT AACATACTGT TTTTTCTTAC TCCACACAGG CATAGAGTGT 2640

CTGCTATTAA TAACTATGCT CAAAAATTGT GTACCTTTAG CTTTTTAATT TGTAAAGGGG 2700

TTAATAAGGA ATATTTGATG TATAGTGCCT TGACTAGAGA TCATAATCAG CCATACCACA 2760

TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC TCCCCCTGAA CCTGAAACAT 2820

AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG CTTATAATGG TTACAAATAA 2880

AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT 2940

TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC GGCTGTGGAA TGTGTGTCAG 3000

TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC 3060

AATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC CAGCAGGCAG AAGTATGCAA 3120

AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC 3180

CTAACTCCGC CCAGTTCCGC CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT 3240

GCAGAGGCCG AGGCCGCCTC GGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT 3300

GGAGGCCTAG GCTTTTGCAA AAAGCTTCAC GCTGCCGCAA GCACTCAGGG CGCAAGGGCT 3360

GCTAAAGGAA GCGGAACACG TAGAAAGCCA GTCCGCAGAA ACGGTGCTGA CCCCGGATGA 3420

ATGTCAGCTA CTGGGCTATC TGGACAAGGG AAAACGCAAG CGCAAAGAGA AAGCAGGTAG 3480

CTTGCAGTGG GCTTACATGG CGATAGCTAG ACTGGGCGGT TTTATGGACA GCAAGCGAAC 3540

CGGAATTGCC AGCTGGGGCG CCCTCTGGTA AGGTTGGGAA GCCCTGCAAA GTAAACTGGA 3600

TGGCTTTCTT GCCGCCAAGG ATCTGATGGC GCAGGGGATC AAGATCTGAT CAAGAGACAG 3660

GATGAGGATC GTTTCGCATG ATTGAACAAG ATGGATTGCA CGCAGGTTCT CCGGCCGCTT 3720

GGGTGGAGAG GCTATTCGGC TATGACTGGG CACAACAGAC AATCGGCTGC TCTGATGCCG 3780

CCGTGTTCCG GCTGTCAGCG CAGGGGCGCC CGGTTCTTTT TGTCAAGACC GACCTGTCCG 3840

GTGCCCTGAA TGAACTGCAG GACGAGGCAG CGCGGCTATC GTGGCTGGCC ACGACGGGCG 3900

TTCCTTGCGC AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG 3960

GCGAAGTGCC GGGGCAGGAT CTCCTGTCAT CTCACCTTGC TCCTGCCGAG AAAGTATCCA 4020

TCATGGCTGA TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC 4080

ACCAAGCGAA ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC 4140

AGGATGATCT GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTC GCCAGGCTCA 4200

AGGCGCGCAT GCCCGACGGC GAGGATCTCG TCGTGACCCA TGGCGATGCC TGCTTGCCGA 4260

ATATCATGGT GGAAAATGGC CGCTTTTCTG GATTCATCGA CTGTGGCCGG CTGGGTGTGG 4320

CGGACCGCTA TCAGGACATA GCGTTGGCTA CCCGTGATAT TGCTGAAGAG CTTGGCGGCG 4380

AATGGGCTGA CCGCTTCCTC GTGCTTTACG GTATCGCCGC TCCCGATTCG CAGCGCATCG 4440

CCTTCTATCG CCTTCTTGAC GAGTTCTTCT GAGCGGGACT CTGGGGTTCG AAATGACCGA 4500

CCAAGCGACG CCCAACCTGC CATCACGAGA TTTCGATTCC ACCGCCGCCT TCTATGAAAG 4560

GTTGGGCTTC GGAATCGTTT TCCGGGACGC CGGCTGGATG ATCCTCCAGC GCGGGGATCT 4620

CATGCTGGAG TTCTTCGCCC ACCCCGGGCT CGATCCCCTC GCGAGTTGGT TCAGCTGCTG 4680

CCTGAGGCTG GACGACCTCG CGGAGTTCTA CCGGCAGTGC AAATCCGTCG GCATCCAGGA 4740

AACCAGCAGC GGCTATCCGC GCATCCATGC CCCCGAACTG CAGGAGTGGG GAGGCACGAT 4800

GGCCGCTTTG GTCCCGGATC TTTGTGAAGG AACCTTACTT CTGTGGTGTG ACATAATTGG 4860

ACAAACTACC TACAGAGATT TAAAGCTCTA AGGTAAATAT AAAATTTTTA AGTGTATAAT 4920

GTGTTAAACT ACTGATTCTA ATTGTTTGTG TATTTTAGAT TCCAACCTAT GGAACTGATG 4980

AATGGGAGCA GTGGTGGAAT GCCTTTAATG AGGAAAACCT GTTTTGCTCA GAAGAAATGC 5040

CATCTAGTGA TGATGAGGCT ACTGCTGACT CTCAACATTC TACTCCTCCA AAAAAGAAGA 5100

GAAAGGTAGA AGACCCCAAG GACTTTCCTT CAGAATTGCT AAGTTTTTTG AGTCATGCTG 5160

TGTTTAGTAA TAGAACTCTT GCTTGCTTTG CTATTTACAC CACAAAGGAA AAAGCTGCAC 5220

TGCTATACAA GAAAATTATG GAAAAATATT CTGTAACCTT TATAAGTAGG CATAACAGTT 5280

ATAATCATAA CATACTGTTT TTTCTTACTC CACACAGGCA TAGAGTGTCT GCTATTAATA 5340

ACTATGCTCA AAAATTGTGT ACCTTTAGCT TTTTAATTTG TAAAGGGGTT AATAAGGAAT 5400

ATTTGATGTA TAGTGCCTTG ACTAGAGATC ATAATCAGCC ATACCACATT TGTAGAGGTT 5460

TTACTTGCTT TAAAAAACCT CCCACACCTC CCCCTGAACC TGAAACATAA AATGAATGCA 5520

ATTGTTGTTG TTAACTTGTT TATTGCAGCT TATAATGGTT ACAAATAAAG CAATAGCATC 5580

ACAAATTTCA CAAATAAAGC ATTTTTTTCA CTGCATTCTA GTTGTGGTTT GTCCAAACTC 5640

ATCAATGTAT CTTATCATGT CTGGATCCCC AGGAAGCTCC TCTGTGTCCT CATAAACCCT 5700

AACCTCCTCT ACTTGAGAGG ACATTCCAAT CATAGGCTGC CCATCCACCC TCTGTGTCCT 5760

CCTGTTAATT AGGTCACTTA ACAAAAAGGA AATTGGGTAG GGGTTTTTCA CAGACCGCTT 5820

TCTAAGGGTA ATTTTAAAAT ATCTGGGAAG TCCCTTCCAC TGCTGTGTTC CAGAAGTGTT 5880

GGTAAACAGC CCACAAATGT CAACAGCAGA AACATACAAG CTGTCAGCTT TGCACAAGGG 5940

CCCAACACCC TGCTCATCAA GAAGCACTGT GGTTGCTGTG TTAGTAATGT GCAAAACAGG 6000

AGGCACATTT TCCCCACCTG TGTAGGTTCC AAAATATCTA GTGTTTTCAT TTTTACTTGG 6060

ATCAGGAACC CAGCACTCCA CTGGATAAGC ATTATCCTTA TCCAAAACAG CCTTGTGGTC 6120

AGTGTTCATC TGCTGACTGT CAACTGTAGC ATTTTTTGGG GTTACAGTTT GAGCAGGATA 6180

TTTGGTCCTG TAGTTTGCTA ACACACCCTG CAGCTCCAAA GGTTCCCCAC CAACAGCAAA 6240

AAAATGAAAA TTTGACCCTT GAATGGGTTT TCCAGCACCA TTTTCATGAG TTTTTTGTGT 6300

CCCTGAATGC AAGTTTAACA TAGCAGTTAC CCCAATAACC TCAGTTTTAA CAGTAACAGC 6360

TTCCCACATC AAAATATTTC CACAGGTTAA GTCCTCATTT AAATTAGGCA AAGGAATTAT 6420

ACACTCCGCT ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG 6480

CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG 6540

TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGGCAGC 6600

GGATCATAAT CAGCCATACC ACATTTGTAG AGGTTTTACT TGCTTTAAAA AACCTCCCAC 6660

ACCTCCCCCT GAACCTGAAA CATAAAATGA ATGCAATTGT TGTTGTTAAC TTGTTTATTG 6720

CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT 6780

TTTCACTGCA TTCTAGTTGT GGTTTGTCCA AACTCATCAA TGTATCTTAT CATGTCTGGA 6840

TCATAATCAG CCATACCACA TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC 6900

TCCCCCTGAA CCTGAAACAT AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG 6960

CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT 7020

CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC 7080

CAGGCCAGAC GCCAACAAGG TAGGAGCTGG AGCATTCGGG CTGGGTTTCA CCCCACCGCA 7140

CGGAGGCCTT TTGGGGTGGA GCCCTCAGGC TCAGGGCATA CTACAAACTT TGCCAGCAAA 7200

TCCGCCTCCT GCCTCCACCA ATCGCCAGTC AGGAAGGCAG CCTACCCCGC TGTCTCCACC 7260

TTTGAGAAAC ACTCATCCTC AGGCCATGCA GTGGAATTCC ACAACCTTCC ACCAAACTCT 7320

GCAAGATCCC AGAGTGAGAG GCCTGTATTT CCCTGCTGGT GGCTCCAGTT CAGGAACAGT 7380

AAACCCTGTT CTGACTACTG CCTCTCCCTT ATCGTCAATC TTCTCGAAAT TCCAGCTGGC 7440

ATTCCGGTAC TGTTGGTAAA ATGGAAGACG CCAAAAACAT AAAGAAAGGC CCGGCGCCAT 7500

TCTATCCTCT AGAGGATGGA ACCGCTGGAG AGCAACTGCA TAAGGCTATG AAGAGATACG 7560

CCCTGGTTCC TGGAACAATT GCTTTTACAG ATGCACATAT CGAGGTGAAC ATCACGTACG 7620

CGGAATACTT CGAAATGTCC GTTCGGTTGG CAGAAGCTAT GAAACGATAT GGGCTGAATA 7680

CAAATCACAG AATCGTCGTA TGCAGTGAAA ACTCTCTTCA ATTCTTTATG CCGGTGTTGG 7740

GCGCGTTATT TATCGGAGTT GCAGTTGCGC CCGCGAACGA CATTTATAAT GAACGTGAAT 7800

TGCTCAACAG TATGAACATT TCGCAGCCTA CCGTAGTGTT TGTTTCCAAA AAGGGGTTGC 7860

AAAAAATTTT GAACGTGCAA AAAAAATTAC CAATAATCCA GAAAATTATT ATCATGGATT 7920

CTAAAACGGA TTACCAGGGA TTTCAGTCGA TGTACACGTT CGTCACATCT CATCTACCTC 7980

CCGGTTTTAA TGAATACGAT TTTGTACCAG AGTCCTTTGA TCGTGACAAA ACAATTGCAC 8040

TGATAATGAA TTCCTCTGGA TCTACTGGGT TACCTAAGGG TGTGGCCCTT CCGCATAGAA 8100

CTGCCTGCGT CAGATTCTCG CATGCCAGAG ATCCTATTTT TGGCAATCAA ATCATTCCGG 8160

ATACTGCGAT TTTAAGTGTT GTTCCATTCC ATCACGGTTT TGGAATGTTT ACTACACTCG 8220

GATATTTGAT ATGTGGATTT CGAGTCGTCT TAATGTATAG ATTTGAAGAA GAGCTGTTTT 8280

TACGATCCCT TCAGGATTAC AAAATTCAAA GTGCGTTGCT AGTACCAACC CTATTTTCAT 8340

TCTTCGCCAA AAGCACTCTG ATTGACAAAT ACGATTTATC TAATTTACAC GAAATTGCTT 8400

CTGGGGGCGC ACCTCTTTCG AAAGAAGTCG GGGAAGCGGT TGCAAAACGC TTCCATCTTC 8460

CAGGGATACG ACAAGGATAT GGGCTCACTG AGACTACATC AGCTATTCTG ATTACACCCG 8520

AGGGGGATGA TAAACCGGGC GCGGTCGGTA AAGTTGTTCC ATTTTTTGAA GCGAAGGTTG 8580

TGGATCTGGA TACCGGGAAA ACGCTGGGCG TTAATCAGAG AGGCGAATTA TGTGTCAGAG 8640

GACCTATGAT TATGTCCGGT TATGTAAACA ATCCGGAAGC GACCAACGCC TTGATTGACA 8700

AGGATGGATG GCTACATTCT GGAGACATAG CTTACTGGGA CGAAGACGAA CACTTCTTCA 8760

TAGTTGACCG CTTGAAGTCT TTAATTAAAT ACAAAGGATA TCAGGTGGCC CCCGCTGAAT 8820

TGGAATCGAT ATTGTTACAA CACCCCAACA TCTTCGACGC GGGCGTGGCA GGTCTTCCCG 8880

ACGATGACGC CGGTGAACTT CCCGCCGCCG TTGTTGTTTT GGAGCACGGA AAGACGATGA 8940

CGGAAAAAGA GATCGTGGAT TACGTCGCCA GTCAAGTAAC AACCGCGAAA AAGTTGCGCG 9000

GAGGAGTTGT GTTTGTGGAC GAAGTACCGA AAGGTCTTAC CGGAAAACTC GACGCAAGAA 9060

AAATCAGAGA GATCCTCATA AAGGCCAAGA AGGGCGGAAA GTCCAAATTG TAAAATGTAA 9120

CTGTATTCAG CGATGACGAA ATTCTTAGCT ATTGTAATGA CTCTAGAGGA TCTTTGTGAA 9180

GGAACCTTAC TTCTGTGGTG TGACATAATT GGACAAACTA CCTACAGAGA TTTAAAGCTC 9240

TAAGGTAAAT ATAAAATTTT TAAGTGTATA ATGTGTTAAA CTACTGATTC TAATTGTTTG 9300

TGTATTTTAG ATTCCAACCT ATGGAACTGA TGAATGGGAG CAGTGGTGGA ATGCCTTTAA 9360

TGAGGAAAAC CTGTTTTGCT CAGAAGAAAT GCCATCTAGT GATGATGAGG CTACTGCTGA 9420

CTCTCAACAT TCTACTCCTC CAAAAAAGAA GAGAAAGGTA GAAGACCCCA AGGACTTTCC 9480

TTCAGAATTG CTAAGTTTTT TGAGTCATGC TGTGTTTAGT AATAGAACTC TTGCTTGCTT 9540

TGCTATTTAC ACCACAAAGG AAAAAGCTGC ACTGCTATAC AAGAAAATTA TGGAAAAATA 9600

TTCTGTAACC TTTATAAGTA GGCATAACAG TTATAATCAT AACATACTGT TTTTTCTTAC 9660

TCCACACAGG CATAGAGTGT CTGCTATTAA TAACTATGCT CAAAAATTGT GTACCTTTAG 9720

CTTTTTAATT TGTAAAGGGG TTAATAAGGA ATATTTGATG TATAGTGCCT TGACTAGAGA 9780

TCATAATCAG CCATACCACA TTTGTAGAGG TTTTACTTGC TTTAAAAAAC CTCCCACACC 9840

TCCCCCTGAA CCTGAAACAT AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG 9900

CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT 9960

CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCC 10020

CCAGGAAGCT CCTCTGTGTC CTCATAAACC CTAACCTCCT CTACTTGAGA GGACATTCCA 10080

ATCATAGGCT GCCCATCCAC CCTCTGTGTC CTCCTGTTAA TTAGGTCACT TAACAAAAAG 10140

GAAATTGGGT AGGGGTTTTT CACAGACCGC TTTCTAAGGG TAATTTTAAA ATATCTGGGA 10200

AGTCCCTTCC ACTGCTGTGT TCCAGAAGTG TTGGTAAACA GCCCACAAAT GTCAACAGCA 10260

GAAACATACA AGCTGTCAGC TTTGCACAAG GGCCCAACAC CCTGCTCAGC AAGAAGCACT 10320

GTGGTTGCTG TGTTAGTAAT GTGCAAAACA GGAGGCACAT TTTCCCCACC TGTGTAGGTT 10380

CCAAAATATC TAGTGTTTTC ATTTTTACTT GGATCAGGAA CCCAGCACTC CACTGGATAA 10440

GCATTATCCT TATCCAAAAC AGCCTTGTGG TCAGTGTTCA TCTGCTGACT GTCAACTGTA 10500

GCATTTTTTG GGGTTACAGT TTGAGCAGGA TATTTGGTCC TGTAGTTTGC TAACACACCC 10560

TGCAGCTCCA AAGGTTCCCC ACCAACAGCA AAAAAATGAA AATTTGACCC TTGAATGGGT 10620

TTTCCAGCAC CATTTTCATG AGTTTTTTGT GTCCCTGAAT GCAAGTTTAA CATAGCAGTT 10680

ACCCCAATAA CCTCAGTTTT AACAGTAACA GCTTCCCACA TCAAAATATT TCCACAGGTT 10740

AAGTCCTCAT TTAAATTAGG CAAAGGAA 10768

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6464 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TTCTTGAAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT 60

AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 120

TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 180

GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT 240

TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT 300

AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG 360

CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA 420

AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG 480

CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT 540

TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC 600

TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA 660

CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT 720

ACCAAACGAC GAGCGTGACA CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAACT 780

ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC 840

GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA 900

TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG 960

TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG 1020

AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA 1080

AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA 1140

GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA 1200

CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG 1260

CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA 1320

TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA 1380

TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC 1440

TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG 1500

TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC 1560

GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT 1620

ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC 1680

GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG 1740

GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG 1800

CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT 1860

GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA 1920

TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG 1980

CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA 2040

TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC 2100

ATAGTTAAGC CAGTATACAC TCCGCTATCG CTACGTGACT GGGTCATGGC TGCGCCCCGA 2160

CACCCGCCAA CACCCGCTGA CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC 2220

AGACAAGCTG TGACCGTCTC CGGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG 2280

AAACGCGCGA GGCAGCGGAT CATAATCAGC CATACCACAT TTGTAGAGGT TTTACTTGCT 2340

TTAAAAAACC TCCCACACCT CCCCCTGAAC CTGAAACATA AAATGAATGC AATTGTTGTT 2400

GTTAACTTGT TTATTGCAGC TTATAATGGT TACAAATAAA GCAATAGCAT CACAAATTTC 2460

ACAAATAAAG CATTTTTTTC ACTGCATTCT AGTTGTGGTT TGTCCAAACT CATCAATGTA 2520

TCTTATCATG TCTGGATCAT AATCAGCCAT ACCACATTTG TAGAGGTTTT ACTTGCTTTA 2580

AAAAACCTCC CACACCTCCC CCTGAACCTG AAACATAAAA TGAATGCAAT TGTTGTTGTT 2640

AACTTGTTTA TTGCAGCTTA TAATGGTTAC AAATAAAGCA ATAGCATCAC AAATTTCACA 2700

AATAAAGCAT TTTTTTCACT GCATTCTAGT TGTGGTTTGT CCAAACTCAT CAATGTATCT 2760

TATCATGTCT GGATCCCAGG CCAGACGCCA ACAAGGTAGG AGCTGGAGCA TTCGGGCTGG 2820

GTTTCACCCC ACCGCACGGA GGCCTTTTGG GGTGGAGCCC TCAGGCTCAG GGCATACTAC 2880

AAACTTTGCC AGCAAATCCG CCTCCTGCCT CCACCAATCG CCAGTCAGGA AGGCAGCCTA 2940

CCCCGCTGTC TCCACCTTTG AGAAACACTC ATCCTCAGGC CATGCAGTGG AATTCCACAA 3000

CCTTCCACCA AACTCTGCAA GATCCCAGAG TGAGAGGCCT GTATTTCCCT GCTGGTGGCT 3060

CCAGTTCAGG AACAGTAAAC CCTGTTCTGA CTACTGCCTC TCCCTTATCG TCAATCTTCT 3120

CGAAATTCCA GCTGGCATTC CGGTACTGTT GGTAAAATGG AAGACGCCAA AAACATAAAG 3180

AAAGGCCCGG CGCCATTCTA TCCTCTAGAG GATGGAACCG CTGGAGAGCA ACTGCATAAG 3240

GCTATGAAGA GATACGCCCT GGTTCCTGGA ACAATTGCTT TTACAGATGC ACATATCGAG 3300

GTGAACATCA CGTACGCGGA ATACTTCGAA ATGTCCGTTC GGTTGGCAGA AGCTATGAAA 3360

CGATATGGGC TGAATACAAA TCACAGAATC GTCGTATGCA GTGAAAACTC TCTTCAATTC 3420

TTTATGCCGG TGTTGGGCGC GTTATTTATC GGAGTTGCAG TTGCGCCCGC GAACGACATT 3480

TATAATGAAC GTGAATTGCT CAACAGTATG AACATTTCGC AGCCTACCGT AGTGTTTGTT 3540

TCCAAAAAGG GGTTGCAAAA AATTTTGAAC GTGCAAAAAA AATTACCAAT AATCCAGAAA 3600

ATTATTATCA TGGATTCTAA AACGGATTAC CAGGGATTTC AGTCGATGTA CACGTTCGTC 3660

ACATCTCATC TACCTCCCGG TTTTAATGAA TACGATTTTG TACCAGAGTC CTTTGATCGT 3720

GACAAAACAA TTGCACTGAT AATGAATTCC TCTGGATCTA CTGGGTTACC TAAGGGTGTG 3780

GCCCTTCCGC ATAGAACTGC CTGCGTCAGA TTCTCGCATG CCAGAGATCC TATTTTTGGC 3840

AATCAAATCA TTCCGGATAC TGCGATTTTA AGTGTTGTTC CATTCCATCA CGGTTTTGGA 3900

ATGTTTACTA CACTCGGATA TTTGATATGT GGATTTCGAG TCGTCTTAAT GTATAGATTT 3960

GAAGAAGAGC TGTTTTTACG ATCCCTTCAG GATTACAAAA TTCAAAGTGC GTTGCTAGTA 4020

CCAACCCTAT TTTCATTCTT CGCCAAAAGC ACTCTGATTG ACAAATACGA TTTATCTAAT 4080

TTACACGAAA TTGCTTCTGG GGGCGCACCT CTTTCGAAAG AAGTCGGGGA AGCGGTTGCA 4140

AAACGCTTCC ATCTTCCAGG GATACGACAA GGATATGGGC TCACTGAGAC TACATCAGCT 4200

ATTCTGATTA CACCCGAGGG GGATGATAAA CCGGGCGCGG TCGGTAAAGT TGTTCCATTT 4260

TTTGAAGCGA AGGTTGTGGA TCTGGATACC GGGAAAACGC TGGGCGTTAA TCAGAGAGGC 4320

GAATTATGTG TCAGAGGACC TATGATTATG TCCGGTTATG TAAACAATCC GGAAGCGACC 4380

AACGCCTTGA TTGACAAGGA TGGATGGCTA CATTCTGGAG ACATAGCTTA CTGGGACGAA 4440

GACGAACACT TCTTCATAGT TGACCGCTTG AAGTCTTTAA TTAAATACAA AGGATATCAG 4500

GTGGCCCCCG CTGAATTGGA ATCGATATTG TTACAACACC CCAACATCTT CGACGCGGGC 4560

GTGGCAGGTC TTCCCGACGA TGACGCCGGT GAACTTCCCG CCGCCGTTGT TGTTTTGGAG 4620

CACGGAAAGA CGATGACGGA AAAAGAGATC GTGGATTACG TCGCCAGTCA AGTAACAACC 4680

GCGAAAAAGT TGCGCGGAGG AGTTGTGTTT GTGGACGAAG TACCGAAAGG TCTTACCGGA 4740

AAACTCGACG CAAGAAAAAT CAGAGAGATC CTCATAAAGG CCAAGAAGGG CGGAAAGTCC 4800

AAATTGTAAA ATGTAACTGT ATTCAGCGAT GACGAAATTC TTAGCTATTG TAATGACTCT 4860

AGAGGATCTT TGTGAAGGAA CCTTACTTCT GTGGTGTGAC ATAATTGGAC AAACTACCTA 4920

CAGAGATTTA AAGCTCTAAG GTAAATATAA AATTTTTAAG TGTATAATGT GTTAAACTAC 4980

TGATTCTAAT TGTTTGTGTA TTTTAGATTC CAACCTATGG AACTGATGAA TGGGAGCAGT 5040

GGTGGAATGC CTTTAATGAG GAAAACCTGT TTTGCTCAGA AGAAATGCCA TCTAGTGATG 5100

ATGAGGCTAC TGCTGACTCT CAACATTCTA CTCCTCCAAA AAAGAAGAGA AAGGTAGAAG 5160

ACCCCAAGGA CTTTCCTTCA GAATTGCTAA GTTTTTTGAG TCATGCTGTG TTTAGTAATA 5220

GAACTCTTGC TTGCTTTGCT ATTTACACCA CAAAGGAAAA AGCTGCACTG CTATACAAGA 5280

AAATTATGGA AAAATATTCT GTAACCTTTA TAAGTAGGCA TAACAGTTAT AATCATAACA 5340

TACTGTTTTT TCTTACTCCA CACAGGCATA GAGTGTCTGC TATTAATAAC TATGCTCAAA 5400

AATTGTGTAC CTTTAGCTTT TTAATTTGTA AAGGGGTTAA TAAGGAATAT TTGATGTATA 5460

GTGCCTTGAC TAGAGATCAT AATCAGCCAT ACCACATTTG TAGAGGTTTT ACTTGCTTTA 5520

AAAAACCTCC CACACCTCCC CCTGAACCTG AAACATAAAA TGAATGCAAT TGTTGTTGTT 5580

AACTTGTTTA TTGCAGCTTA TAATGGTTAC AAATAAAGCA ATAGCATCAC AAATTTCACA 5640

AATAAAGCAT TTTTTTCACT GCATTCTAGT TGTGGTTTGT CCAAACTCAT CAATGTATCT 5700

TATCATGTCT GGATCCCCAG GAAGCTCCTC TGTGTCCTCA TAAACCCTAA CCTCCTCTAC 5760

TTGAGAGGAC ATTCCAATCA TAGGCTGCCC ATCCACCCTC TGTGTCCTCC TGTTAATTAG 5820

GTCACTTAAC AAAAAGGAAA TTGGGTAGGG GTTTTTCACA GACCGCTTTC TAAGGGTAAT 5880

TTTAAAATAT CTGGGAAGTC CCTTCCACTG CTGTGTTCCA GAAGTGTTGG TAAACAGCCC 5940

ACAAATGTCA ACAGCAGAAA CATACAAGCT GTCAGCTTTG CACAAGGGCC CAACACCCTG 6000

CTCAGCAAGA AGCACTGTGG TTGCTGTGTT AGTAATGTGC AAAACAGGAG GCACATTTTC 6060

CCCACCTGTG TAGGTTCCAA AATATCTAGT GTTTTCATTT TTACTTGGAT CAGGAACCCA 6120

GCACTCCACT GGATAAGCAT TATCCTTATC CAAAACAGCC TTGTGGTCAG TGTTCATCTG 6180

CTGACTGTCA ACTGTAGCAT TTTTTGGGGT TACAGTTTGA GCAGGATATT TGGTCCTGTA 6240

GTTTGCTAAC ACACCCTGCA GCTCCAAAGG TTCCCCACCA ACAGCAAAAA AATGAAAATT 6300

TGACCCTTGA ATGGGTTTTC CAGCACCATT TTCATGAGTT TTTTGTGTCC CTGAATGCAA 6360

GTTTAACATA GCAGTTACCC CAATAACCTC AGTTTTAACA GTAACAGCTT CCCACATCAA 6420

AATATTTCCA CAGGTTAAGT CCTCATTTAA ATTAGGCAAA GGAA 6464

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TGASTCA 7

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TGGNNNNNNN GCCCAA 16

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TGGCA 5

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TGACACA 7

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TGAGTCA 7

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TGANACA 7

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TGATACA 7

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CCNTGTNT 8

WE CLAIM:

1. A method for quantifying the amount of transforming growth factor-ß (TGF-ß) in a liquid sample, which method comprises:

(a) incubating said liquid sample together with eucaryotic cells that contain a TGF-ß responsive expression vector having a gene encoding luciferase for a predetermined time period sufficient for said eucaryotic cells to express a detectable amount of said luciferase;

(b) measuring the amount of said luciferase expressed during said time period; and

(c) determining the amount of TGF-ß present in said sample by comparing the measured amount of said luciferase against a reference curve.

2. The method in accordance with claim 1 wherein the reference curve represents a series of measured amounts of said luciferase produced from a series of known concentrations of TGF-ß by said eucaryotic cells.

3. The method in accordance with claim 1 wherein said eucaryotic cells are mammalian cells.

4. The method in accordance with claim 3 wherein said mammalian cells are members of the group consisting of mink lung epithelial cells, HeLa cells, Chinese hamster ovary cells, Hep3B cells, GM7373 cells, and NIH 3T3 cells.

5. The method in accordance with claim 1 wherein the TGF-ß responsive expression vector is a plasmid comprising, in the direction of transcription, a regulatory region that includes at least one TGF-ß inducible response element that is operatively linked to a promoter, and a structural region downstream of said promoter, said response element being capable of inducing dose-dependent luciferase activity and said structural region coding for said luciferase.

6. The method in accordance with claim 5 wherein said plasmid includes a nucleotide sequence that corresponds to a

sequence selected from the group consisting of SEQ ID NOs 1-10.

7. The method in accordance with claim 5 wherein said plasmid has the identifying characteristics of a plasmid selected from the group consisting of plasmid ATCC Accession Number 75627, plasmid ATCC Accession Number 74628 and plasmid ATCC Accession Number 75629.

8. The method in accordance with claim 5 wherein said TGF-ß inducible response element comprises a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 11-17.

9. The method in accordance with claim 5 wherein said promoter comprises a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 18 and 19.

10. The method in accordance with claim 1 wherein said eucaryotic cells are stably transformed cells that contain said TGF-ß responsive vector, and wherein said vector also includes a gene encoding a selectable marker.

11. The method in accordance with claim 10 wherein said vector is a plasmid comprising a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 1-6.

12. The method in accordance with claim 1 wherein said eucaryotic cells are transiently transformed cells that contain said TGF-ß responsive vector, and wherein said vector is a plasmid comprising a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 7-10.

13. The method in accordance with claim 1 wherein said liquid sample is selected from the group consisting of a body fluid, culture medium and a tissue extract.

14. A method for quantifying the amount of transforming growth factor-ß (TGF-ß) in a liquid sample comprising:

(a) providing, in eucaryotic cells capable of expressing an indicator molecule, a plasmid comprising, in the direction of transcription, a regulatory region that includes

at least one TGF-ß inducible response element that is operatively linked to a promoter, and a structural region downstream of said promoter, said response element being capable of inducing dose-dependent indicator molecule activity and said structural region coding for said indicator molecule;

(b) incubating said liquid sample with said eucaryotic cells for a predetermined time period sufficient for said eucaryotic cells to express a detectable amount of said indicator molecule;

(c) measuring the amount of said indicator molecule expressed during said time period; and

(d) comparing the measured amount of said indicator molecule produced in step (c) with the amount of indicator molecule produced in a control assay performed according to steps (a) through (c) by treating said liquid sample with an anti-TGF-ß antibody to obtain a net measured amount of said indicator molecule induced by said TGF-ß.

15. The method in accordance with claim 14 wherein said liquid sample contains an isoform of TGF-ß selected from the group consisting of TGF-ß1, TGF-ß2 and TGF-ß3.

16. The method in accordance with claim 14 wherein said liquid sample is selected from the group consisting of a body fluid, culture medium and a tissue extract. 17. The method in accordance with claim 14 wherein said eucaryotic cell is a mammalian cell.

18. The method in accordance with claim 14 wherein said mammalian cell is selected from the group consisting of mink lung epithelial cells, HeLa cells, Chinese Hamster Ovary cells, Hep3B cells, GM7373 cells and NIH 3T3 cells.

19. The method in accordance with claim 14 wherein said indicator molecule is luciferase.

20. The method in accordance with claim 14 wherein said plasmid comprises a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 1-10.

21. The method in accordance with claim 14 wherein said

TGF-ß inducible response element comprises a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 11-17.

22. The method in accordance with claim 14 wherein said promoter comprises a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 18 and 19.

23. The method in accordance with claim 14 wherein said plasmid has the identifying characteristics of a plasmid selected from the group consisting of plasmid ATCC Accession Number 75627, plasmid ATCC Accession Number 74628 and plasmid ATCC Accession Number 75629.

24. The method in accordance with claim 14 wherein said eucaryotic cells are stably transformed cells that contain said plasmid, and wherein said plasmid contains a gene encoding a selectable marker for the selection of said stably transformed cells.

25. The method in accordance with claim 24 wherein said plasmid comprises a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 1-6.

26. The method in accordance with claim 14 wherein said eucaryotic cells are stably transformed cells that contain the TGF-ß response element having the nucleotide sequence in SEQ ID NO 11, and wherein said cells correspond to cells on deposit with ATCC having the ATCC Accession Number CRL 11508.

27. The method in accordance with claim 14 wherein eucaryotic cells comprise transiently transformed cells that contain said plasmid comprising a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 7-10.

28. The method in accordance with claim 14 further comprising the step of:

(e) determining the amount of said TGF-ß present in said sample by comparing the measured amount of said indicator molecule obtained in step (d) against a reference curve.

29. The method in accordance with claim 28 wherein said reference curve represents a series of measured amounts of said indicator molecule produced from a series of known concentrations of TGF-ß in said eucaryotic cells.

30. A plasmid vector in substantially pure form capable of causing expression of an indicator molecule in a eucaryotic cell, said plasmid including in the direction of transcription, a first nucleotide sequence comprising a regulatory region that includes at least one TGF-ß inducible response element operatively linked to a promoter, a second nucleotide sequence comprising a structural region downstream of said promoter and coding for said indicator molecule, and a third nucleotide sequence comprising a gene encoding a selectable marker for the selection of a stably transformed cell, said response element being capable of inducing dose-dependent luciferase activity and said structural region coding for said luciferase.

31. The plasmid vector in accordance with claim 30 capable of expressing a chemiluminescent indicator molecule.

32. The plasmid vector in accordance with claim 30 wherein said plasmid comprises a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 1-6.

33. The plasmid vector in accordance with claim 30 wherein said TGF-ß inducible response element comprises a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 11-17.

34. The plasmid vector in accordance with claim 30 wherein said promoter comprises a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 18 and 19.

35. The plasmid vector in accordance with claim 30 wherein said gene comprises the nucleotide sequence in SEQ ID NO 20.

36. A plasmid vector in substantially pure form and capable of causing expression of luciferase in a eucaryotic

cell, said plasmid comprising in the direction of transcription, a regulatory region that includes at least one TGF-ß inducible response element that is operatively linked to a promoter, and a structural region downstream of said promoter for transcription therefrom and coding for said luciferase, said response element being capable of inducing dose-dependent luciferase activity and said structural region coding for said luciferase, and wherein said plasmid has the identifying characteristics of a plasmid selected from the group consisting of plasmid ATCC Accession Number 75627, plasmid ATCC Accession Number 74628 and plasmid ATCC Accession Number 75629.

37. A plasmid vector in substantially pure form and capable of causing expression of luciferase in a eucaryotic cell, said plasmid comprising in the direction of transcription, a regulatory region that includes at least one TGF-ß inducible response element that is operatively linked to a promoter, and a structural region downstream of said promoter for transcription therefrom and coding for said luciferase, said response element being capable of inducing dose-dependent luciferase activity and said structural region coding for said luciferase, and wherein said plasmid comprises a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID Nos 7-10.

38. A eucaryotic cell containing a plasmid vector having a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 1-10.

39. The eucaryotic cell in accordance with claim 38 wherein said cell is selected from the group consisting of mink lung epithelial cells, HeLa cells, Chinese hamster ovary cells, Hep3B cells, GM7373 cells and NIH 3T3 cells.

40. A kit useful in assaying the amount of TGF-ß in a liquid sample comprising (a) packaging material; (b) eucaryotic cells contained within said packaging material, said cells capable of expressing an indicator molecule and containing a plasmid comprising, in the direction of transcription, a

regulatory region that includes at least one TGF-ß inducible response element that is operatively linked to a promoter, and a structural region downstream of said promoter, said response element being capable of inducing dose-dependent indicator molecule activity and said structural region coding for said indicator molecule; and (c) an aliquot of TGF-ß contained within said packaging material, said TGF-ß used for generating a reference curve representing a measured amount of the indicator molecule produced from a known concentration of TGF-ß.

41. The kit in accordance with claim 40 wherein said eucaryotic cells are selected from the group consisting of mink lung epithelial cells, HeLa cells, Chinese Hamster Ovary cells, Hep3B cells, GM7373 cells and NIH 3T3 cells.

42. The kit in accordance with claim 40 wherein said plasmid comprises a nucleotide sequence that corresponds to a sequence selected from the group consisting of SEQ ID NOs 1-10.

43. The kit in accordance with claim 40 wherein said plasmid comprises a plasmid having the identifying characteristics of a plasmid selected from the group consisting of plasmid ATCC Accession Number 75627, plasmid ATCC Accession Number 74628 and plasmid ATCC Accession Number 75629.

44. The kit in accordance with claim 40 wherein said packaging material comprises a label indicating that said eucaryotic cells can be used for determining the amount of TGF-ß in said liquid sample comprising the steps of (a) incubating said cells with said liquid sample; (b) measuring the amount of said indicator molecule produced thereby; and (c) comparing the amount of measured indicator molecule with said reference curve.

45. The kit in accordance with claim 40 wherein said eucaryotic cells are stably transformed cells that contain the TGF-ß response element having the nucleotide sequence in SEQ ID NO 11, and wherein said cells correspond to cells on deposit with ATCC having the ATCC Accession Number CRL 11508.

46. The kit in accordance with claim 40 further comprising: (d) an anti-TGF-ß antibody for use in a parallel control assay for determining the amount of indicator molecule produced other than by TGF-ß induction.

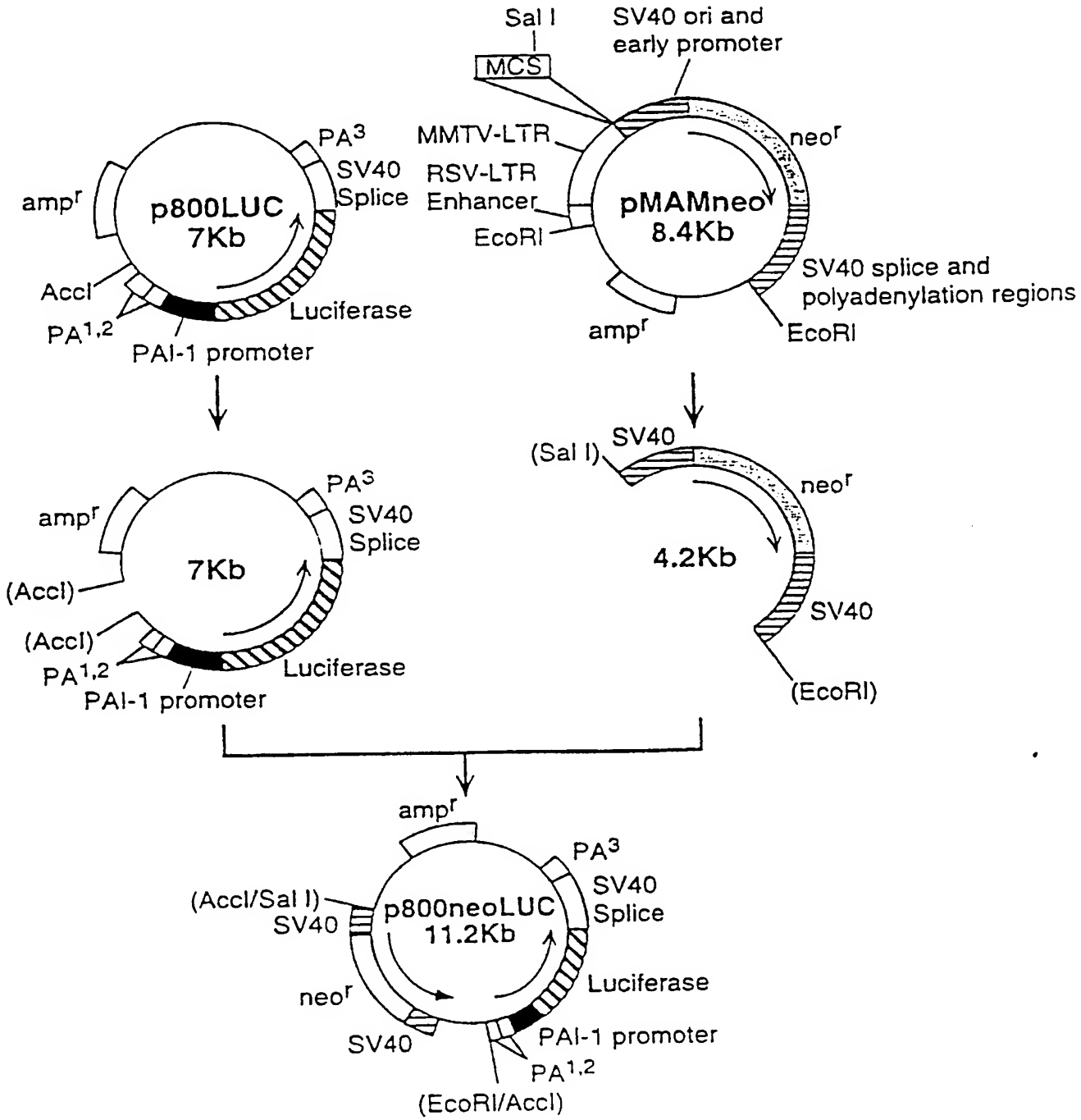
Sal I
SV40 ori and early promoter
MCS
MMTV-LTR
RSV-LTR Enhancer
EcoRI
neor
pMAMneo 8.4Kb
SV40 splice and polyadenylation regions
ampr
EcoRI
PA3
SV40 Splice
ampr
p800LUC 7Kb
AccI
PA1,2
PAI-1 promoter
Luciferase
SV40
(Sal I)
neor
4.2Kb
SV40
(EcoRI)
PA3
SV40 Splice
ampr
7Kb
(AccI)
(AccI)
PA1,2
PAI-1 promoter
Luciferase
ampr
PA3
SV40 Splice
(AccI/Sal I)
SV40
p800neoLUC 11.2Kb
neor
Luciferase
SV40
PAI-1 promoter
PA1,2
(EcoRI/AccI)

FIGURE 1

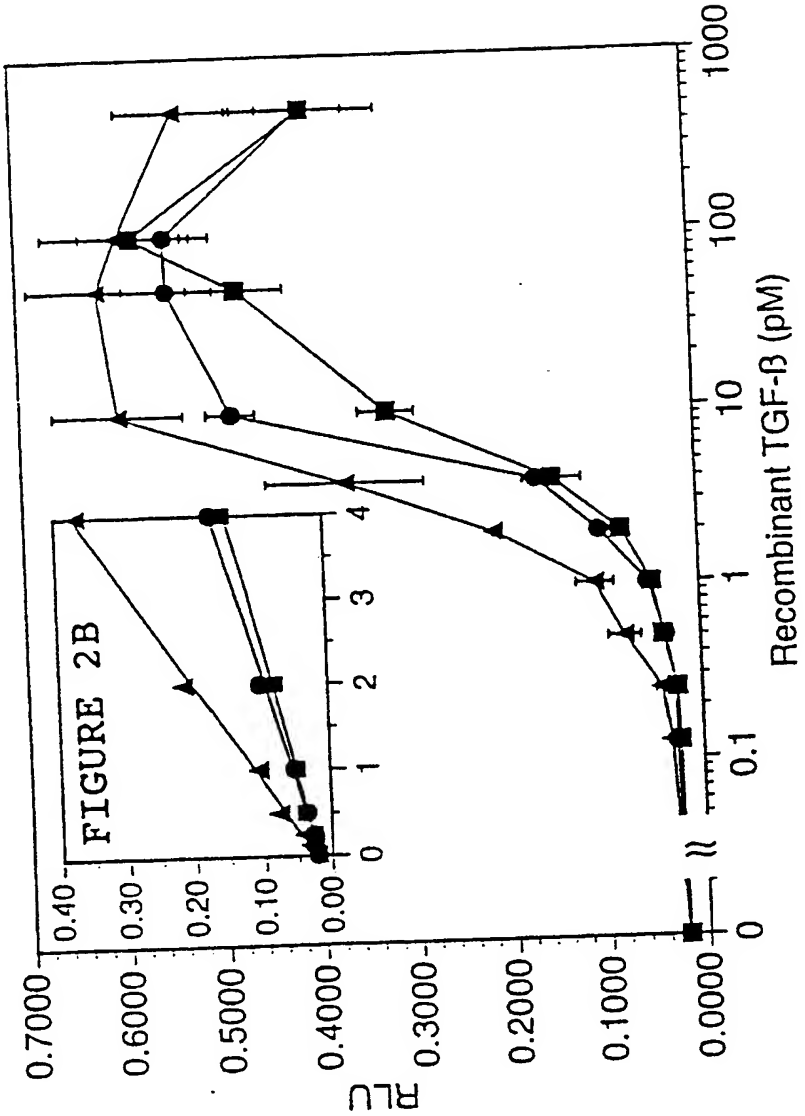
RLU
0.7000
0.6000
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
0
0.1
1
10
100
1000
Recombinant TGF-β (pM)
FIGURE 2B
0.40
0.30
0.20
0.10
0.00
0
1
2
3
4

FIGURE 2A

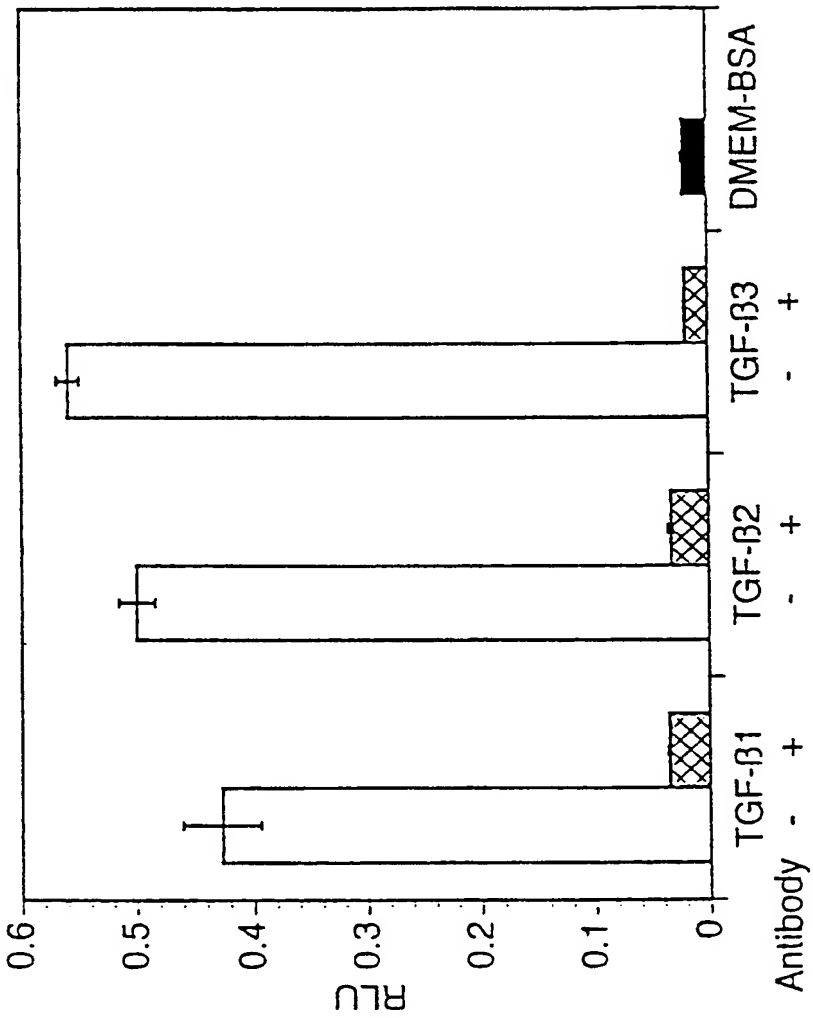

FIGURE 2C

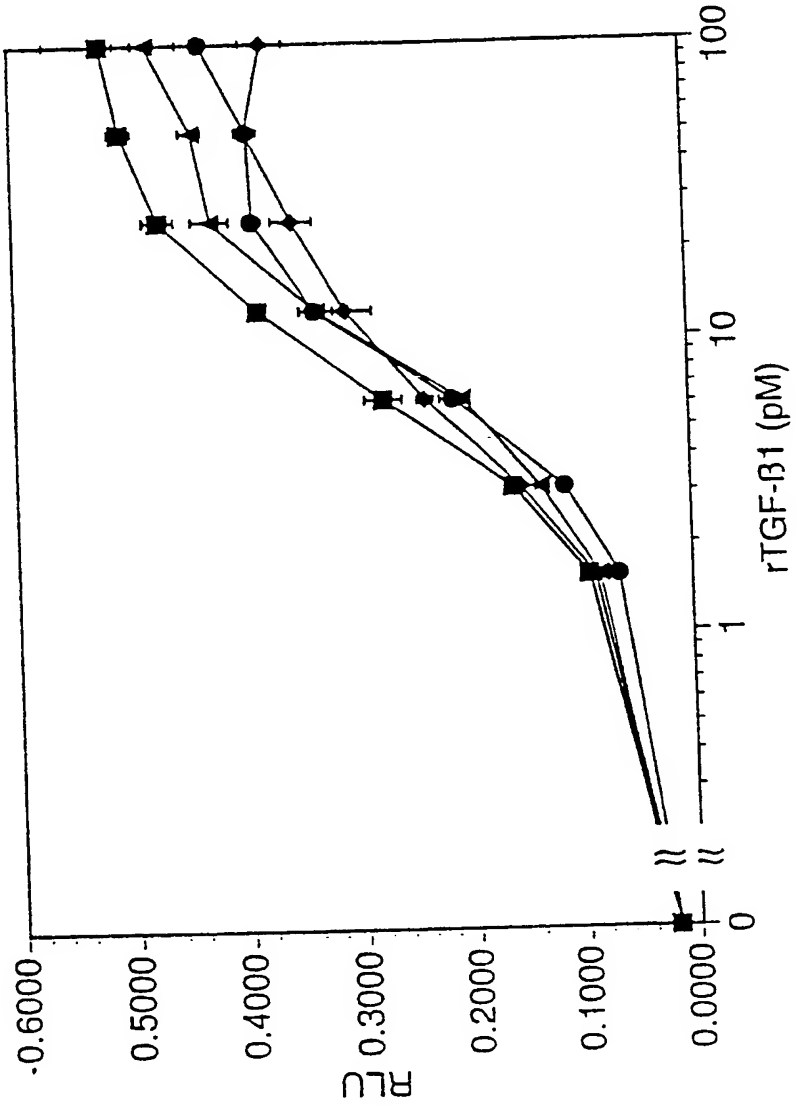
RLU
-0.6000
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
0
1
10
100
rTGF-β1 (pM)

FIGURE 3A

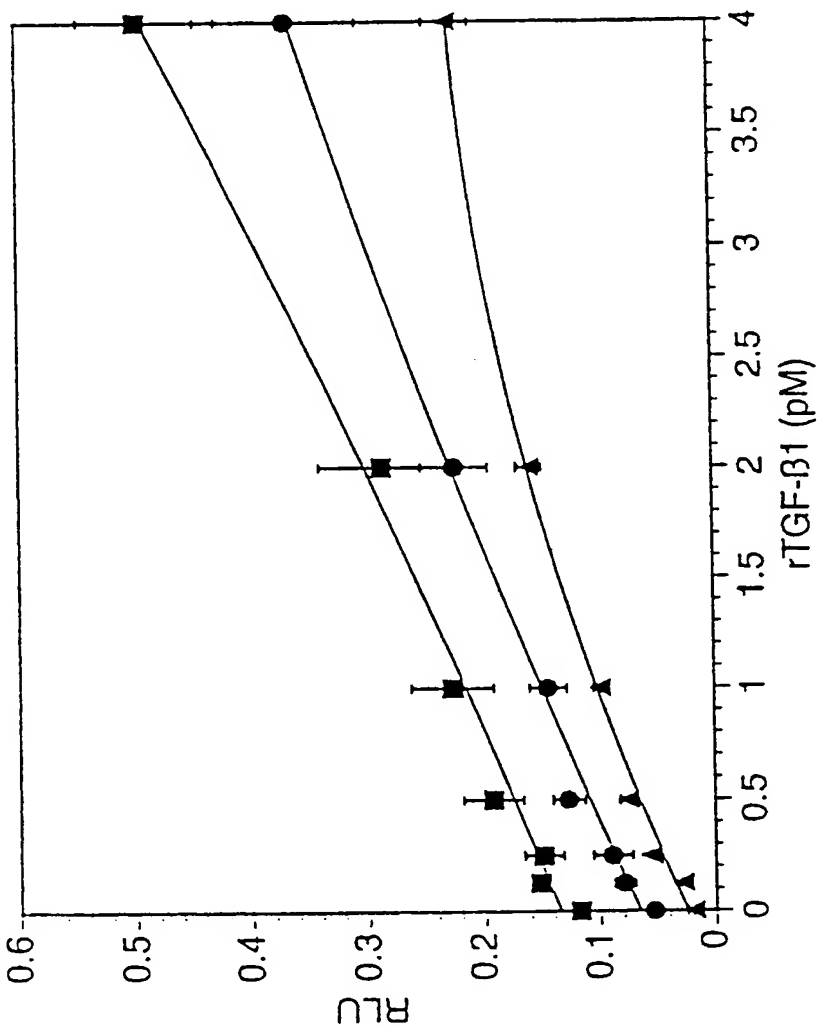

FIGURE 3B

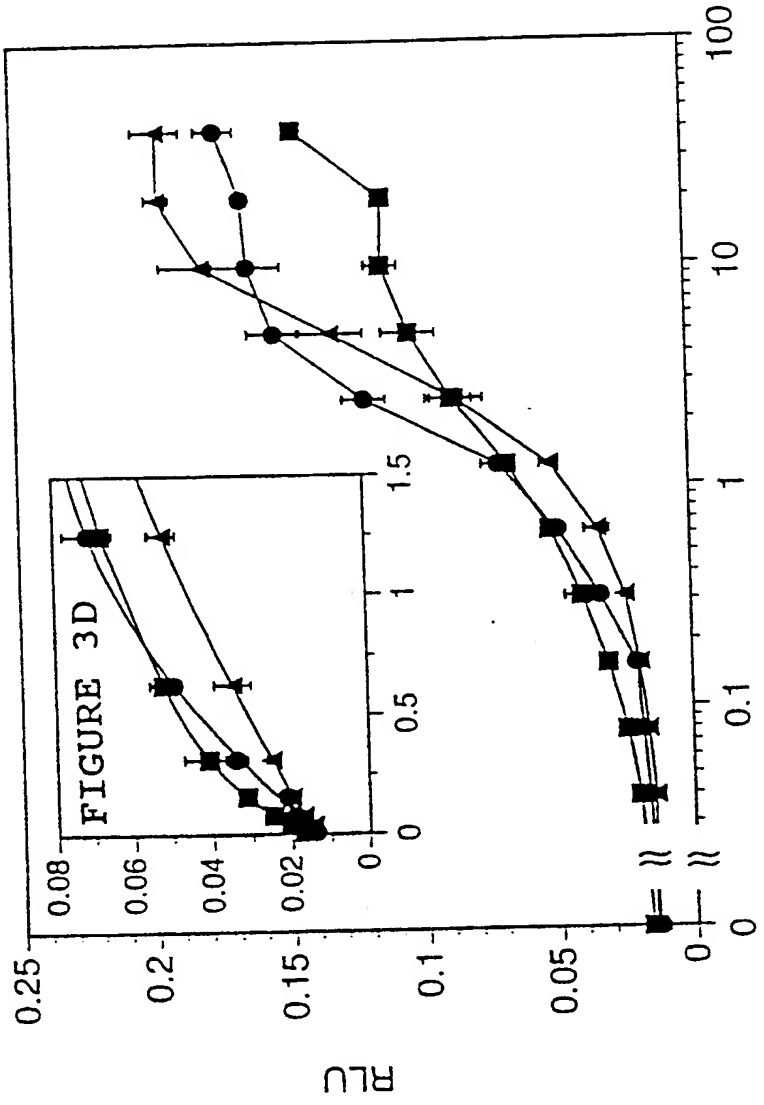
RLU
0.25
0.2
0.15
0.1
0.05
0
0
0.1
1
10
100
TGF-β1 (pM)
FIGURE 3D
0.08
0.06
0.04
0.02
0
0
0.5
1
1.5

FIGURE 3C

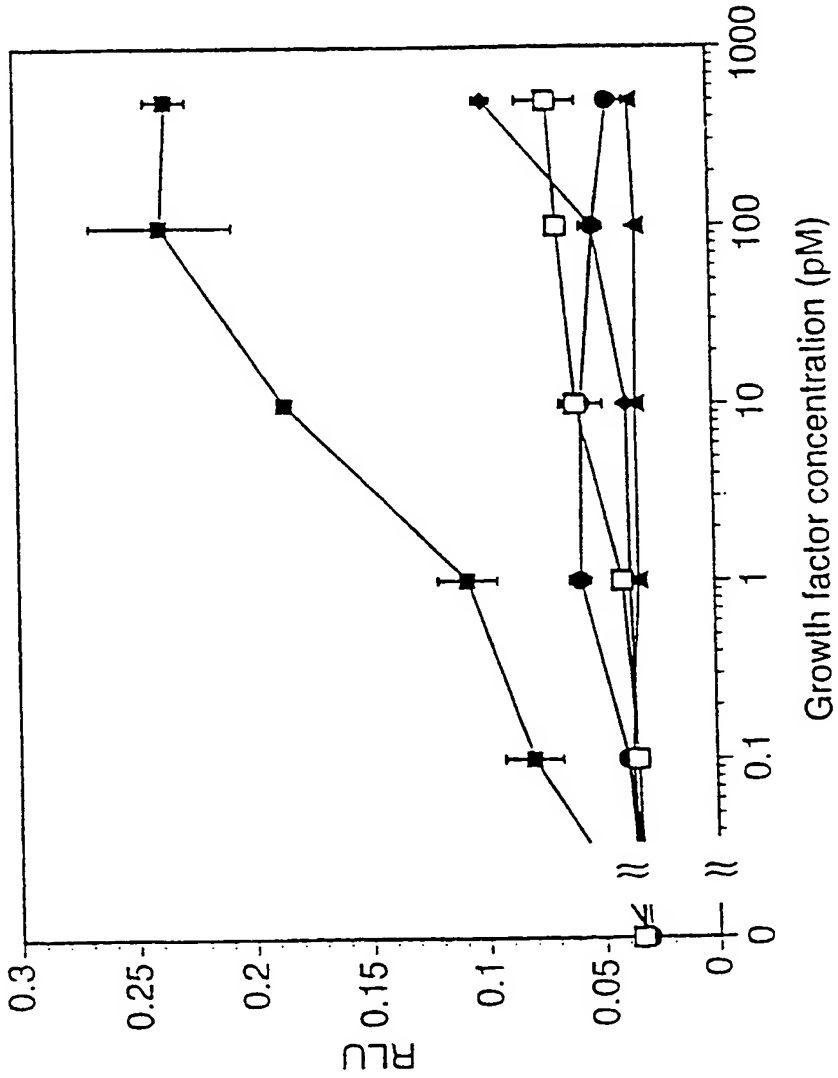
RLU
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
1
10
100
1000
Growth factor concentration (pM)

FIGURE 4A

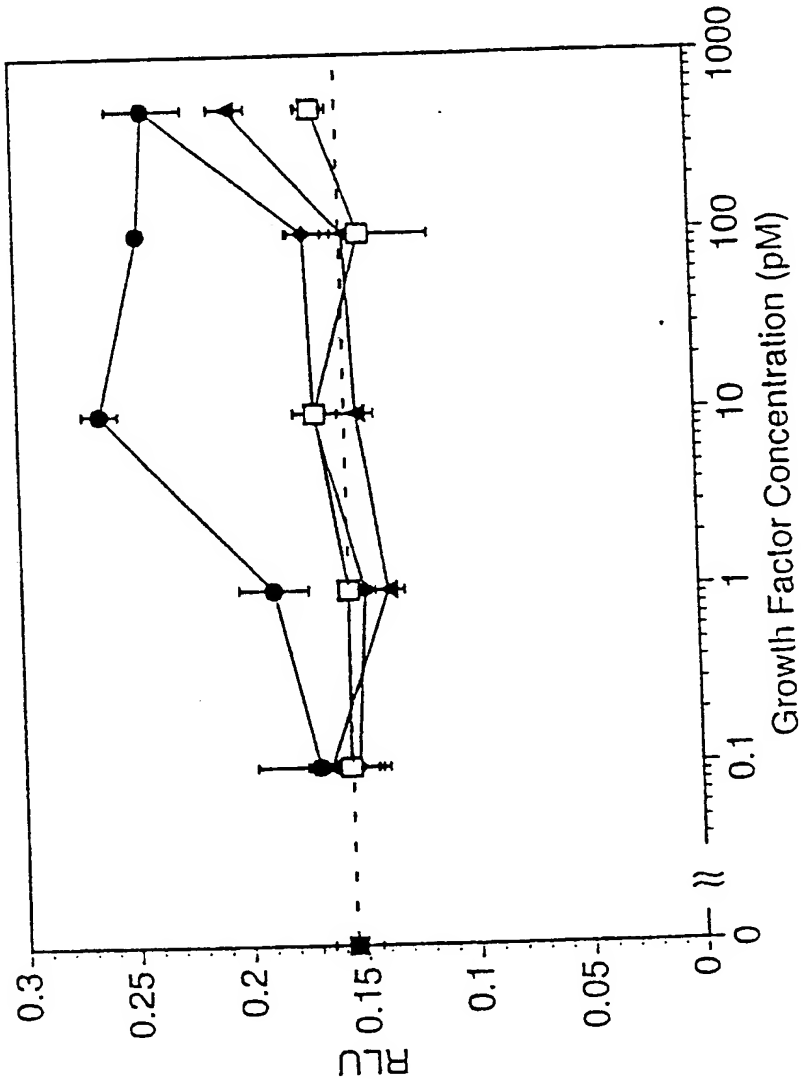
RLU
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
1
10
100
1000
Growth Factor Concentration (pM)

FIGURE 4B

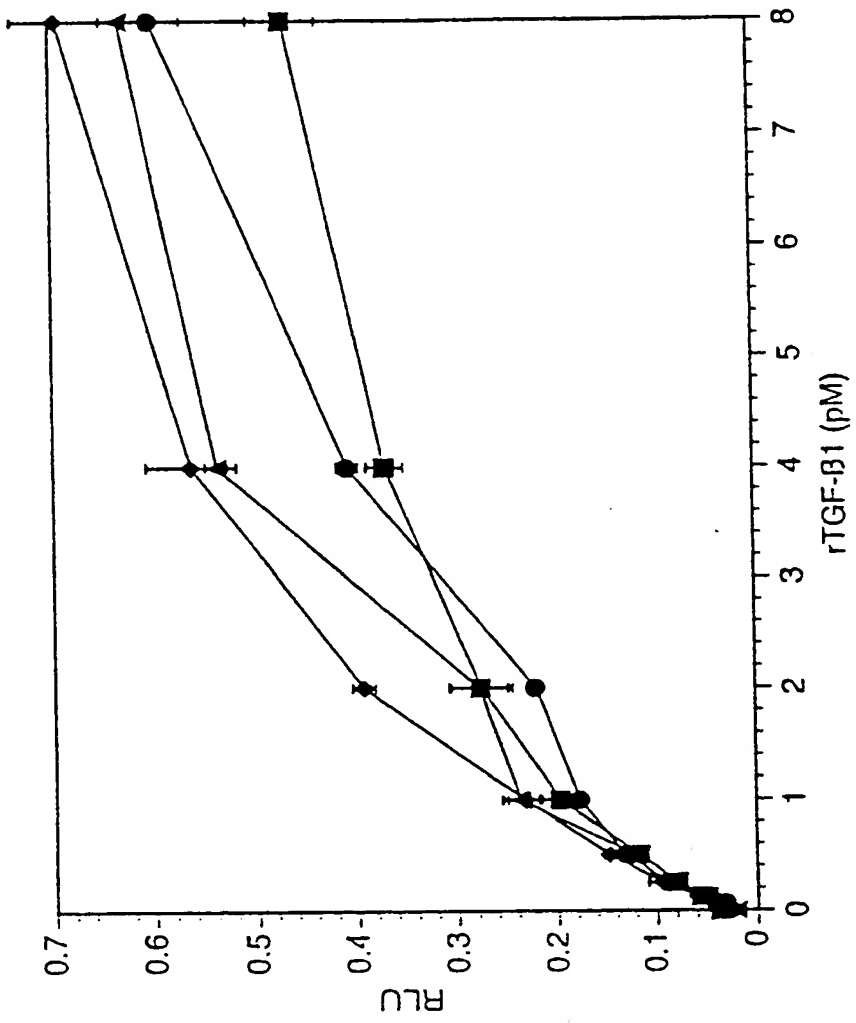

FIGURE 4C

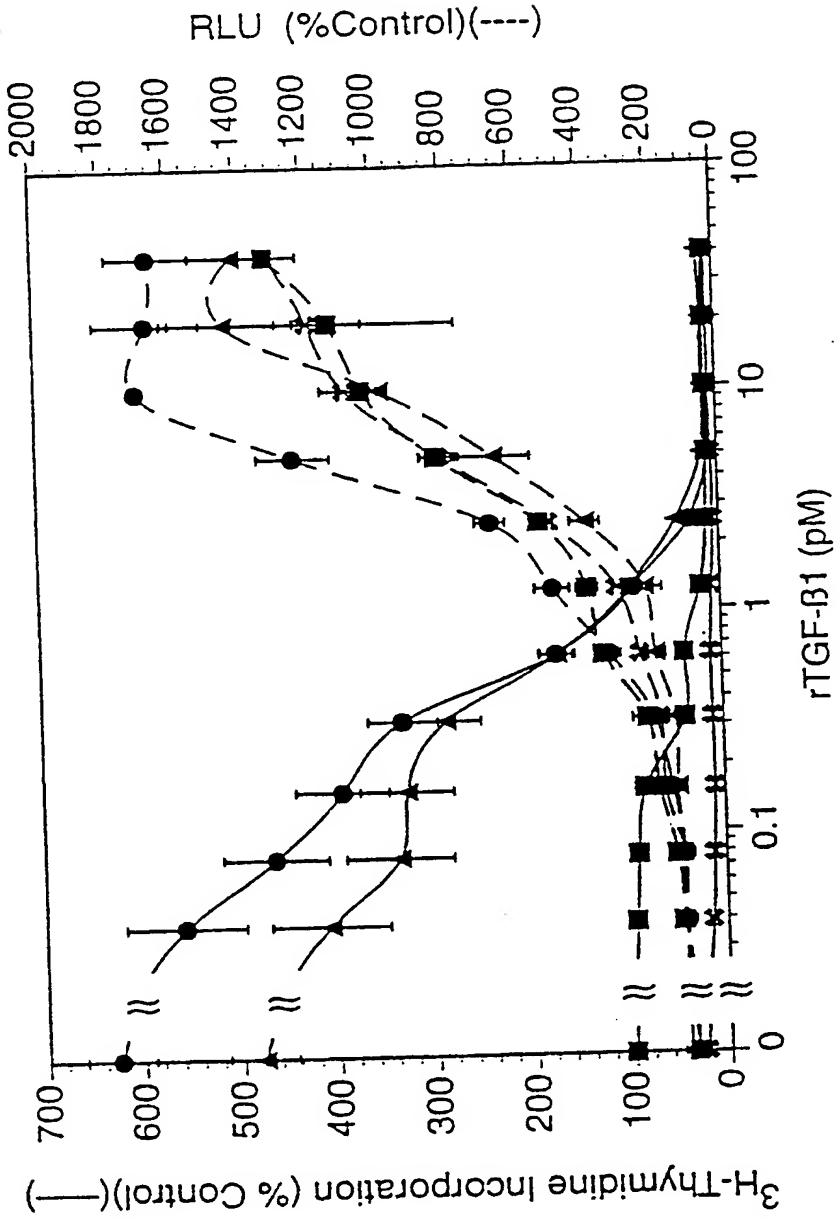
RLU (%Control)(----)
2000
1800
1600
1400
1200
1000
800
600
400
200
0
100
10
1
0.1
0
rTGF-β1 (pM)
700
600
500
400
300
200
100
0
^{3}H-Thymidine Incorporation (% Control)(——)

FIGURE 5

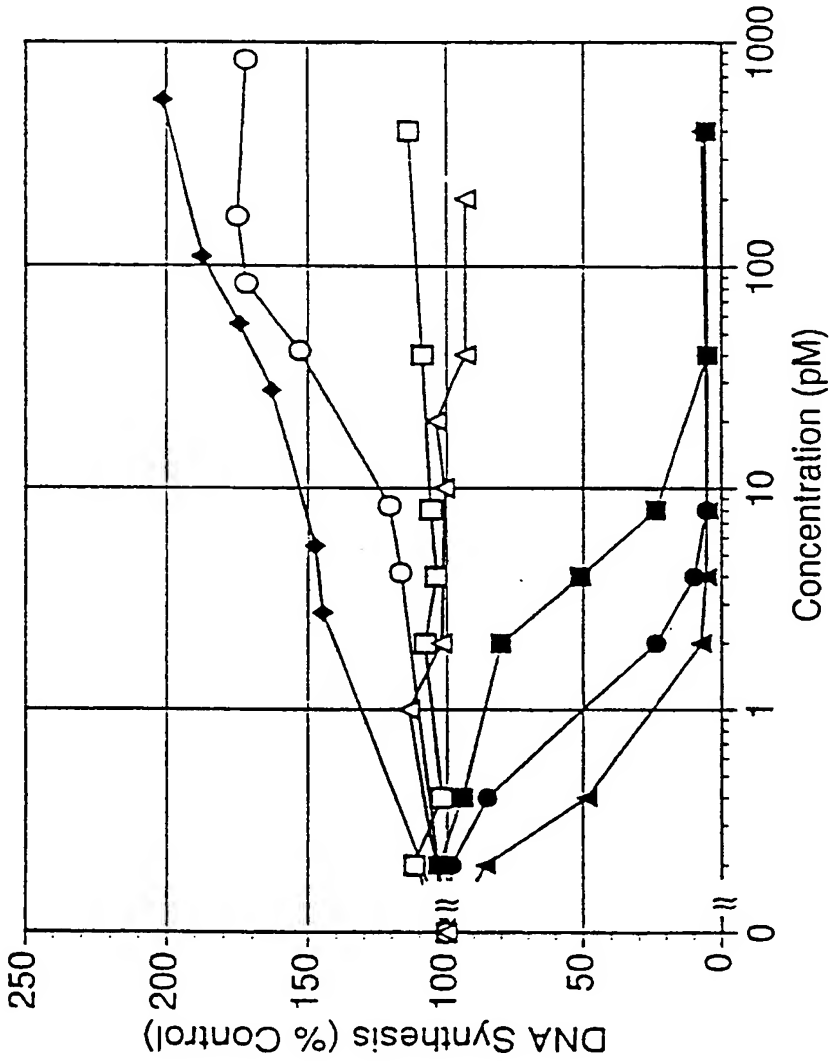
DNA Synthesis (% Control)
250
200
150
100
50
0
Concentration (pM)
0
1
10
100
1000

FIGURE 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/01153

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07H 21/00; C12N 15/18; C07K 14/495
US CL :435/8, 69.1, 69.4, 320.1, 240.1; 530/399

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/8, 69.1, 69.4, 320.1, 240.1; 530/399

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Cell, Volume 71, issued 11 December 1992, J.L. Wrana et al., "TGFβ Signals through a Heteromeric Protein Kinase Receptor Complex," pages 1003-1014, especially figure 1A and Results.	1-46
Y	Journal of Cellular Physiology, Volume 152, issued 1992, R. Flaumenhaft et al., "Cell Density Dependent Effects of TGF-β Demonstrated by a Plasminogen Activator-Based Assay for TGF-β," pages 48-55, especially figure 5.	1-46
Y	The Journal of Biological Chemistry, Volume 193, issued 1951, O.H. Lowry et al., "Protein Measurement with the Folin Phenol Reagent," pages 265-275, especially pages 265-268.	1-46

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 11 APRIL 1995	Date of mailing of the international search report 21APR1995
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Shelly Guest Cermak Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/01153

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US, A, 5,216,126 (COX ET AL.) 01 June 1993.	1-46
A	US, A, 5,268,295 (SERRERO) 07 December 1993.	1-46
A	Molecular and Cellular Biology, Volume 12, No. 4, issued April 1992, A. Riccio et al., "Transforming Growth Factor β1-Responsive Element: Closely Associated Binding Sites for USF and CCAAT-Binding Transcription Factor-Nuclear Factor I in the Type 1 Plasminogen Activator Inhibitor Gene," pages 1846-1855.	1-46

INTERNATIONAL SEARCH REPORT

ernational application No.
PCT/US95/01153

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

Sequence search of PIR, GeneSeq,
APS
Dialog
search terms: transforming growth factor, quantification, luciferase